Technical Report RSC-42

WATER QUALITY PARAMETER MEASUREMENT USING SPECTRAL SIGNATURES

by

P. E. White

May 1973

supported by

National Aeronautics and Space Administration
NASA Grant NGL 44-001-001

TEXAS A&M UNIVERSITY
REMOTE SENSING CENTER
COLLEGE STATION, TEXAS

Technical Report RSC-42

WATER QUALITY PARAMETER MEASUREMENT USING SPECTRAL SIGNATURES

by

P. E. White

May 1973

supported by

National Aeronautics and Space Administration

NASA Grant NGL 44-001-001

ABSTRACT

Increased concern about water quality and the lack of fast, inexpensive methods of measuring water quality parameters over large areas have prompted interest in remote sensing techniques. Previous studies indicate that some water quality parameters influence remotely sensed spectral signatures, but currently there is no method of measuring water quality parameters from spectral signatures.

This study applies regression analysis to the problem of measuring water quality parameters from remote sensing spectral signature data. It presents the equations necessary to perform regression analysis and describes methods of testing the strength and reliability of a regression. It also presents an efficient algorithm for selecting an optimal subset of the independent variables available for a regression.

To illustrate the regression techniques, they are applied to a particular set of data from the Houston Ship Channel. Unfortunately, deficiencies in the quality and quantity of the data that were available prevent definite conclusions to be drawn from the analysis.

Results of the study show that regression analysis is applicable to the problem of measuring water quality parameters from remote sensing spectral signatures, but

that the utility of the method needs further testing on much larger sets of data.

Recommendations resulting from this study are that regression analysis techniques be applied to data from ERTS-1, and that new water quality parameters based on the reflective characteristics of water be defined for improved correlation with remote sensing data.

ACKNOWLEDGEMENTS

Funding for this study was provided by the National Aeronautics and Space Administration (NGL 44-001-001) and the United States Army. Collection and processing of the data in this study would have been impossible without the personnel of the Johnson Space Center who provided the remote sensing data, the personnel of the Environmental Engineering Department of Texas A&M University who provided equipment and assistance for collection of ground observation data, Optronics International Incorporated who densitized the photography, and Richard Carter of the Data Processing Center of Texas A&M who digitized the scanner data. Assistance in preparing and editing the manuscript were provided by Dr. W. P. James and Dr. J. P. German, with special thanks to Dr. J. W. Rouse, Jr., for his support and guidance. I am indebted to Mrs. Joyce Boyle, Mrs. Karen Zeier, Mrs. Connie Schilhab, Mrs. Marie Bush, and Miss Karen Brown for their typing at various stages in the preparation of this manuscript.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Uncontrolled waste discharges imperil the productivity and aesthetic value of waterways and coastal zones. Almost all industry, agriculture, and cities produce wastes, and dumping waste into waterways is often the most convenient method of disposal. When the levels of waste discharges are low, the water removes the wastes and cleans itself. At higher levels of waste discharge, the capabilities of the water must be determined and waste discharges kept within those capabilities.

A reliable sensor capable of measuring water quality parameters is needed to determine acceptable levels of waste discharges and to monitor water quality. Current *in situ* methods of measuring these parameters are slow, expensive, and often limited by weather and terrain.

Studies indicate that changes in some water quality parameters produce changes in the spectral reflectance signature of water obtained by airborne remote sensors. A method of determining water quality parameters from remote sensing data would greatly improve water quality determina-

---

The citations on the following pages follow the style of the Proceedings of IEEE.

ation. This report will attempt to use regression analysis to develop such a method.

Problem Description

Sensor Applications

One use for an efficient, inexpensive method of monitoring water quality parameters would be modeling hydrological systems. An accurate model would make possible realistic estimates of water quality improvement to justify the expense of changing waste disposal procedures and equipment. Such models would also facilitate placement of waste discharges. Since hydrological systems are non-uniform, proper placement of discharges might ameliorate their adverse effects with little or no additional cost.

A second use for a sensor to monitor water quality parameters would be enforcement of waste discharge regulations. The sensor could measure waste concentrations and flow rates to determine the quantities of waste discharged. The locations of waste discharges would be indicated by concentration gradients.

A third use for a sensor to measure water quality parameters would be determination of general water quality within a system. This would indicate how much improvement was needed and how much had already been achieved. It

would also indicate significant spatiotemporal anomalies.

Sensor Requirements

In general, hydrological systems are large and have many components. For instance, efforts to improve the water quality of a bay must consider not only the bay and any wastes discharged directly into it, but must also consider rivers and streams which also contribute to the wastes in the bay. While the bay may be large, the combined system is even larger. Besides being large, hydrological systems are dynamic. They are influenced by factors such as topography, wind, weather, and tides, and they can change as quickly as these factors can change.

To be effective with large, dynamic hydrological systems, a sensor for water quality parameters must be fast, must provide spatial resolution, and must be accurate. Modeling requires a fast sensor to collect all the required data before changes occur in the system. A fast sensor is also required for enforcement to detect and measure temporary discharges.

Spatial resolution, the capability to distinguish between small areas, is necessary to obtain concentration gradients and to locate isolated regions with unusual conditions. The gradients are useful for developing

system models and for locating and measuring discharges to enforce regulations.

To be useful, a sensor must possess relative and/or absolute accuracy. Relative accuracy, the capability of indicating which regions have higher or lower values of water quality parameters, could indicate concentration variations within the system and could indicate concentration gradients. Absolute accuracy, the capability to give actual concentrations, would provide the information necessary to evaluate water quality and to measure quantities of wastes for enforcement of waste discharge regulations.

Need for Airborne Remote Sensors

The current method of monitoring water quality is *in situ* sampling. Although this method is normally accurate, it is expensive, slow, and has limited resolution. Expenses involve not only sampling devices, but also the personnel to operate them. Speed is limited by the number of sampling devices being used and the time required to move them. Resolution is limited by the number of sampling devices and by the area that they must cover.

The primary characteristics of airborne remote sensors are just those lacked by *in situ* sampling methods.

Airborne sensors can instantly view large areas and can move quickly from one area to another, regardless of terrain. Such sensors provide essentially continuous coverage of the surface and good resolution may be obtained.

Besides the features mentioned above, experience has indicated the possibility that remote sensors can be used to measure some water quality parameters. The passive sensors to be considered measure elements of the spectral reflectance signature of an object. Laboratory measurements show that different aqueous solutions have different signatures. It should then be possible to determine some of the impurities in water by comparing the signatures of polluted water to that of pure water.

Historical Background

The art of remote sensing was greatly advanced when photography was combined with the aerial view. Photography was capable of obtaining more data than the eye alone because it could be made to respond to a wider spectrum of radiation. In addition, photography could hold data for more detailed study or for comparison with other photography. The aerial view made the photography even more useful since it exposed large surface areas

simultaneously. This permitted spatial orientation and facilitated comparison of different subareas within the field of view [1].

An indication of the potential of spectral signatures is the fact that different objects appear with different darknesses in different bands. Colwell [1], discussing the advantages of multiband imagery, gave the simple example in table I-1. This indicates that some features may be distinguished by simultaneous consideration of two bands. For instance, clear water may be distinguished from turbid water, and water may be distinguished from cement. The table also indicates that some features, such as turbid water and macadam, must be distinguished by shape or texture or possibly by using more elements in their spectral signatures.

The delineation of oil spills and of different types of sea water also indicate the potential for using spectral signatures. Estes and Golomb [2] studied the feasibility of using various remote sensors to measure oil spills. They found standard panchromatic minus blue photography and color imagery ineffective, but found thermal infrared imagery useful. In thermal infrared imagery, it seemed that areas with thin layers of oil appeared darker than the surrounding water while areas

TABLE I-1

SOME TWO BAND SIGNATURES

Category and Feature	Multispectral Signature	
	Pan-25A (.58-.72 Band)	IR-89B (.72-.89 Band)
1. Water Bodies		
a. Clear Water	Dark	Dark
b. Turbid Water	Light	Dark
2. Forest Trees		
a. Hardwoods	Dark	Light
b. Conifers	Dark	Dark
3. Improved Roads		
a. Cement	Light	Light
b. Macadam	Light	Dark

with thick oil layers appeared lighter.

Ewing [3] studied variations in sea water spectra as a function of biological activity as evidenced by chlorophyll content, which was often associated with variations in water temperature. Near George Bank, it was found that the two sides of a temperature front with accompanying color variations could consistently be delineated by their spectra at altitudes from 500 to 10,000 feet.

A spectral signature provides some desciption of apparent spectral radiance as a function of wavelength. The theory of obtaining spectral signatures is quite simple. Objects reflect and emit electro-magnetic radiation. To obtain the signature of an object, sensors are used to measure the energy content of one or more bands of the radiated energy. Comparison of energy to the wavelength of the corresponding band, usually in the form of a graph, constitutes a spectral signature [1,4].

In a given band, the fraction of incident radiation an object reflects depends in part on the composition of the object and wavelength of the band. The same is true of radiation emitted by the object. Some substances may be characterized by the bands having high and low apparent radiances. These characteristics gives spectral signatures

their usefulness [1].

When impurities are present in water, the signature of the combination is a composite of their signatures. If the impurities produce characteristic high apparent radiances in bands for which water produces low apparent radiances, or if the reverse is true, the concentration of the impurity may be estimated by comparing the composite signature to that of pure water [4].

Using work done by Yentsch [5] and Ewing [3], an illustration of the use of spectral signatures may be presented. From two different sources, Yentsch obtained the curves for pure water and distilled sea water shown in figure I-1. From cultures of various phytoplankton, he obtained the signatures shown in figure I-2. Figures I-1 (p. 10) and I-2 (p. 11) indicate that water has a minimum attenuation and phytoplankton a maximum attenuation at short wavelengths, and that water has a high attenuation and phytoplankton a low attenuation at long wavelengths. This suggests that as the concentration of phytoplankton increases in water, short wavelengths should be attenuated more and long wavelengths attenuated less. Curves prepared by Ewing and shown in figure I-3 indicate that this is true.

James and Burgess [6] studied the feasibility of measuring waste concentrations of pulp mill outfalls

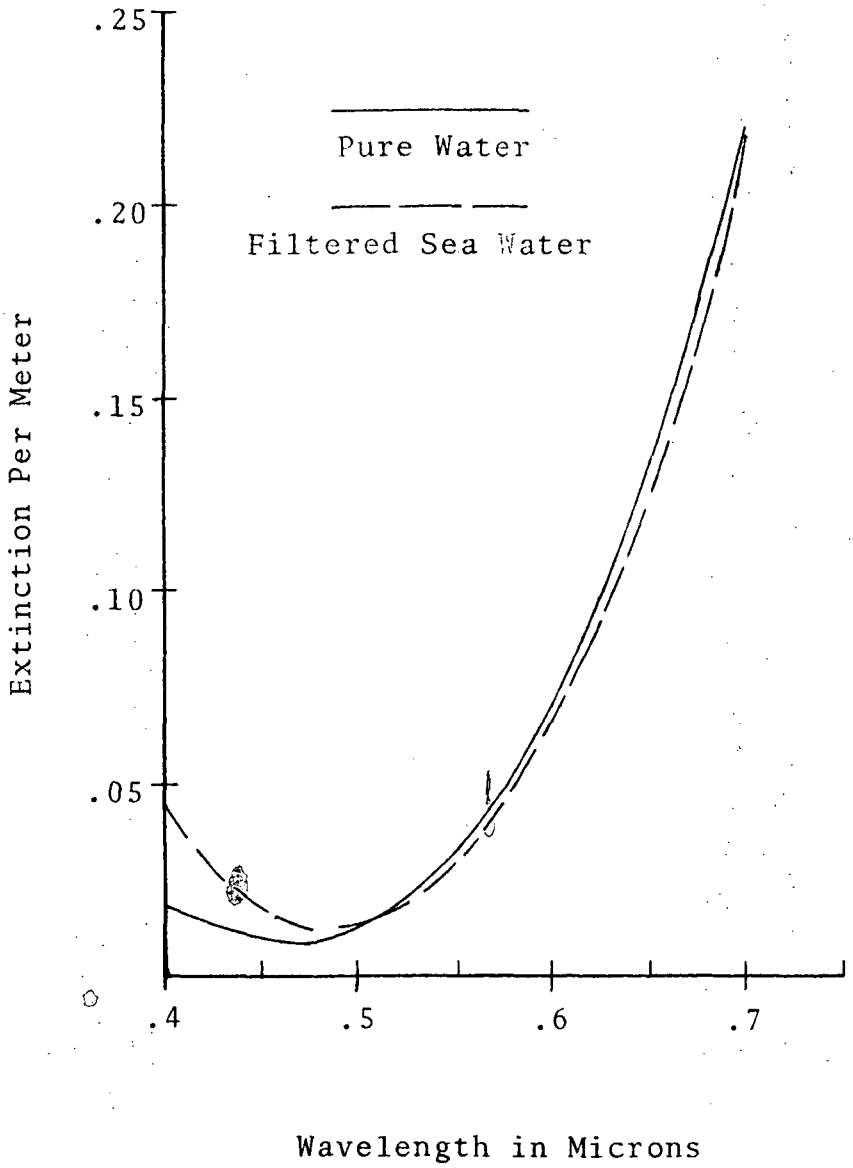


Fig. I-1 Signatures of Pure Water and Filtered Sea Water [5]

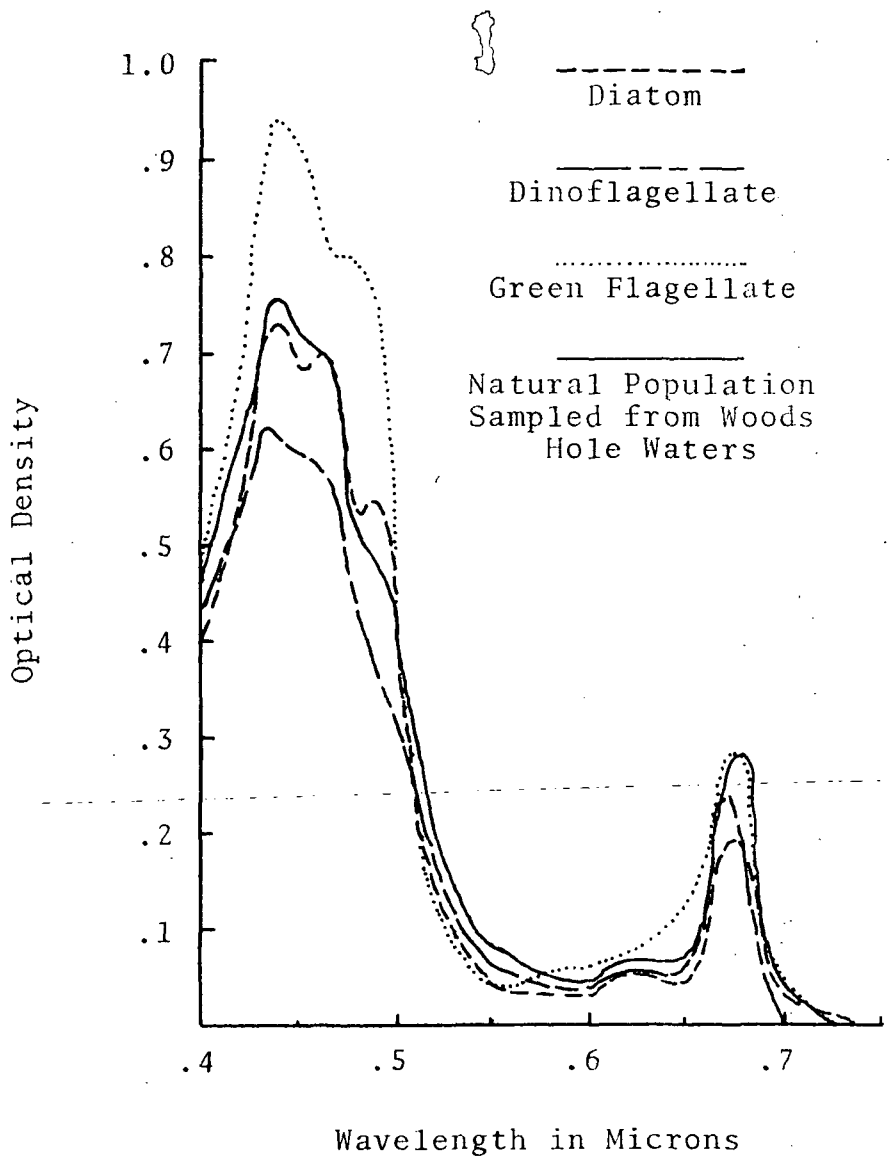


Fig. I-2 Optical Density of Phytoplankton[5]

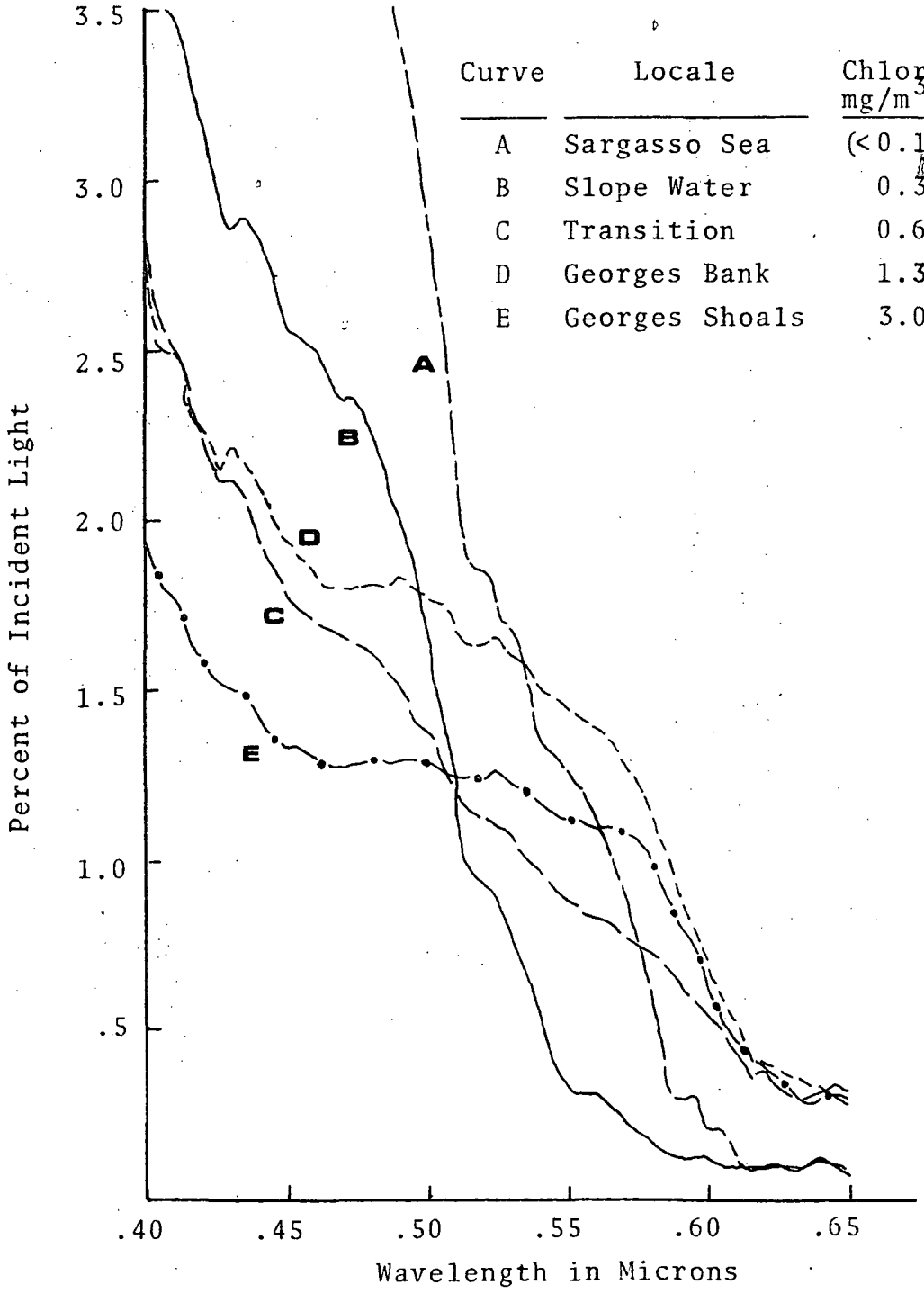


Fig. I-3 Scattered Radiation from Water with Different Concentrations of Chlorophyll (Phytoplankton)[3]

using photographic data. They found that the ratio of red to green light was the photographic variable most sensitive to variations in waste concentration. The correlation coefficient between the photographic data and boat sampling data was about .9.

Report Objectives

The foregoing discussion seeks to make three points. First, some method of measuring water quality parameters is needed. Second, it would be useful to be able to measure water quality parameters using airborne remote sensors. Third, spectral signatures obtained from airborne remote sensors are influenced by at least some water quality parameters. This leaves the problem of how to obtain quantitative values of water quality parameters from the values of the spectral signatures.

Some water quality parameters that are of interest and which might be expected to influence spectral signatures are color, transmittance and turbidity. The data analysis of this report will primarily deal with these parameters.

The relation between values of water quality parameters and values of the elements of the spectral signatures must be established using experimental data. This

suggests the use of regression analysis which is designed to provide such a relationship.

To perform a regression analysis, a relation of the form

$$Y = A_o + \sum_{i=1}^{n} \sum_{j=1}^{m_i} A_{ij} f_{ij}(X_i) \qquad \text{(I-1)}$$

where: Y = the dependent variable

X_i = one of the independent variables

n = the number of independent variables

m_i = the number of different functions of X_i being included in the regression

$f_{ij}(X_i)$ = a function of X_i

A_o = a constant

A_{ij} = a constant coefficient of $f_{ij}(X_i)$

is assumed. The values of the m_i's and the forms of the f_{ij}'s are obtained from theory, by assumption, or by examination of scatter diagrams of dependent versus independent parameters. Using experimental data, estimates of the values of A_o and the A_{ij}'s are then calculated using a least squares criterion.

A technique related to and used with regression analysis is correlation analysis. Using a set of assumptions about the distributions of the data values,

correlation analysis provides a method of verifying that the regression does specify a meaningful relationship between the dependent and independent variables. In addition, it aids in selecting optimal subsets of the terms in the full regression to obtain a less complicated reduced regression.

This report will attempt to do four things. First, it will describe selection of the f_{ij}'s for water quality data. Second, it will describe how to perform the regression to obtain the values of A_o and of the A_{ij}'s. Third, it will describe how to perform the correlation analysis and how to test the regression. Finally, it will describe how to select optimal subsets of the regression terms to obtain simplified regressions.

Scope of Report

Chapter II provides the background for this report. It describes spectral signatures and explains how they may be obtained using photography or multispectral scanner data.

Chapter III describes the data analysis techniques. It includes the equations necessary to perform the regression and correlation, and the algorithm used to facilitate selection of optimal subsets of the regression terms.

Chapter IV describes how data were obtained to illustrate the techniques presented in this report. It discusses the nature and rationale of the experiment, the collection of ground observation data, and the processing of photography and multispectral scanner data prior to analysis.

Chapter V illustrates the application of the data analysis techniques. It gives the results obtained when the analysis techniques were applied to the data described in Chapter IV.

Chapter VI gives the conclusions and recommendations drawn from the development of the analysis and from the analysis of the sample data set.

CHAPTER II

BACKGROUND

The sensors considered in this report include photography and multispectral scanners. This chapter provides a brief description of these sensors and of the types of information they provide.

Spectral Signatures

Emitted Radiation

Photography and multispectral scanners are both passive sensors. The radiation they measure is either reflected or emitted by the targets they detect. At normal temperatures, reflected radiation predominates for wavelengths below 3.5 microns, and emitted radiation predominates for wavelengths above 4 microns [7].

All substances emit radiation. For solids and liquids, the spectrum of the emitted radiation is continuous [8]. Although the actual spectrum depends on temperature, surface roughness, wavelength and properties of the material, the maximum limit is given by the blackbody equation as illustrated by the curves of figure II-1 [9]. These curves indicate that for normal temperatures around 300°K, emitted radiation is primarily at wavelengths well

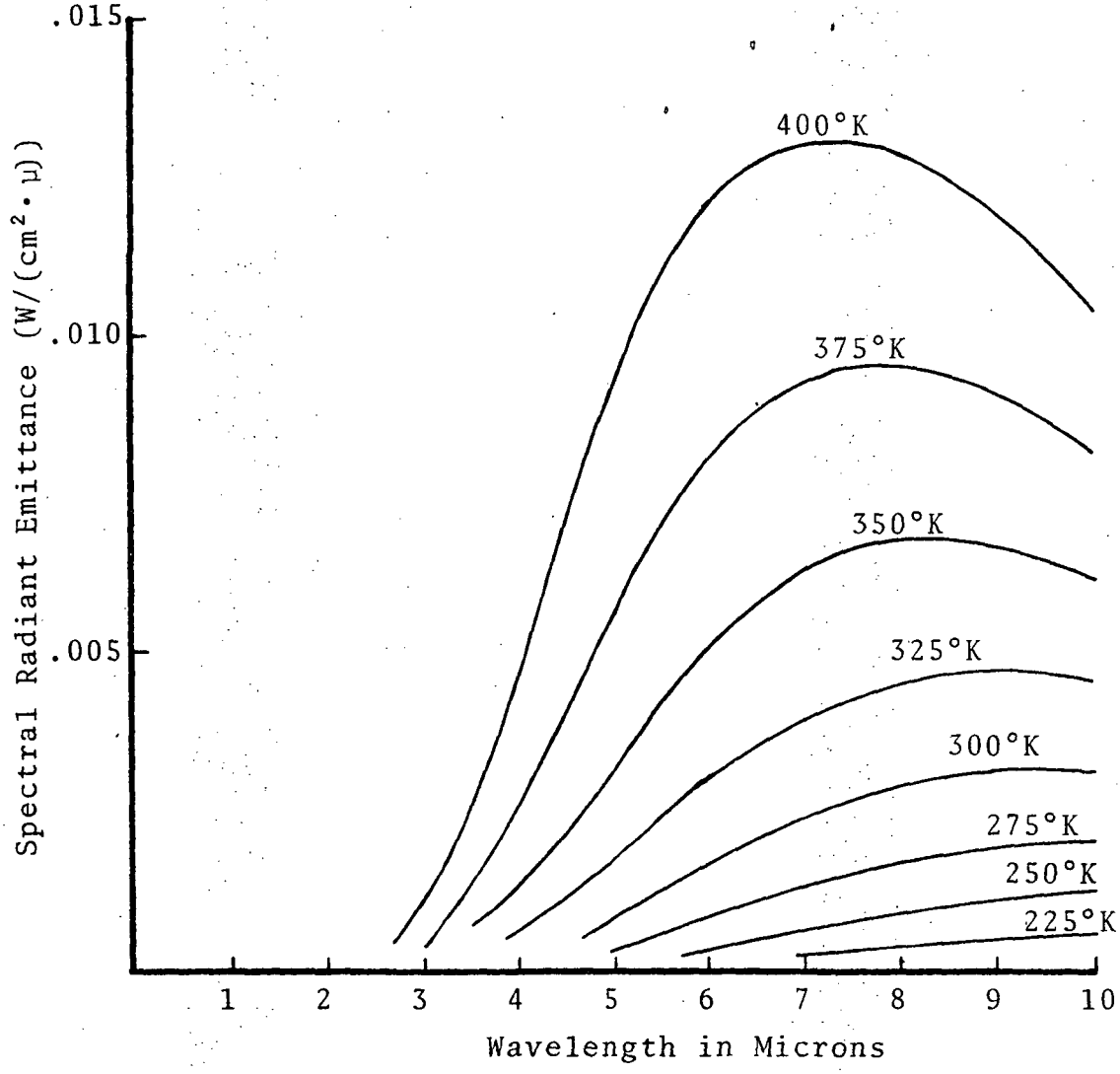


Fig. II-1 Blackbody Radiation Curves

above 4 microns. This is outside the range of interest of this thesis.

Incident Radiation

The main contribution to spectral signatures that is of concern in this thesis is scattered and reflected sunlight. Most water quality parameters depend in some way on waste concentrations. To obtain insight into the relationship between scattered radiation, i.e. spectral signatures, and water quality parameters, the relationship between scattered radiation and waste concentrations will be developed. The development will follow closely the work of James [10].

This development will proceed in two steps: first, the interaction of incident radiation with the water and contaminants will be considered; second, extraneous effects producing noise and interference will be considered. The first step will consider surface effects as the radiation enters and leaves the water, and volume effects as it propagates through the water. The result will be an expression for the scattered radiance immediately above the surface of the water as a function of contaminant concentration.

Radiation-Water Interaction. The geometry for the radiation-water interaction problem is shown in figure II-2.

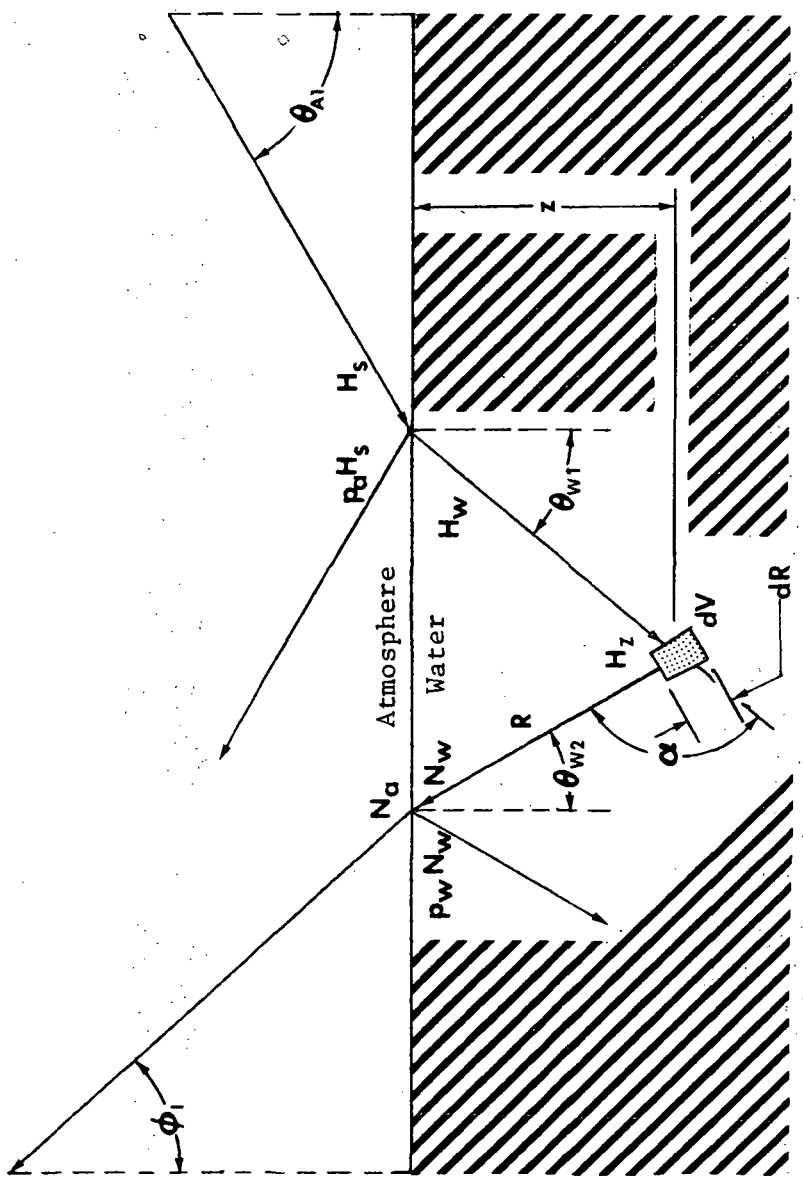


Fig. II-2 Geometry for Radiation-Water Interaction

As the radiation reaches the surface of the water, part is reflected and part is refracted. Assuming an optically flat surface and applying Fresnel's law, the fraction that is reflected is

$$\rho_a = \frac{1}{2}\left[\frac{\tan^2(\theta_{Al} - \theta_{Wl})}{\tan^2(\theta_{Al} + \theta_{Wl})} + \frac{\sin^2(\theta_{Al} - \theta_{Wl})}{\sin^2(\theta_{Al} + \theta_{Wl})}\right] \qquad \text{(II-1)}$$

where: ρ_a = fraction of radiation reflected at air-water interface

θ_{Al} = see figure II-2 (p. 21)

θ_{Wl} = see figure II-2 (p. 21)

The angle of refraction, θ_{Wl}, can be obtained from Snell's law,

$$n = \frac{\sin \theta_{Al}}{\sin \theta_{Wl}} \qquad \text{(II-2)}$$

where: n = index of refraction of water $\approx 4/3$

The subsurface irradiance normal to the beam is given by

$$H_W = H_S \frac{1 - \rho_a}{\cos \theta_{Wl}} \qquad \text{(II-3)}$$

where: H_W = direct irradiance at the upper surface of the water

H_S = direct irradiance at the lower surface of the atmosphere

As the radiation travels through the water, it is

scattered and absorbed by the water and by any contaminants. This results in attenuation of the radiation. Defining an attenuation coefficient, c_A, given by

$$c_A = a + bW \qquad \text{(II-4)}$$

where: c_A = attenuation coefficient
a = water attenuation coefficient
b = contaminant absorption coefficient
W = contaminant concentration

the subsurface irradiance normal to the beam at depth z is given by

$$H_z = H_W \exp(-c_A z \sec \theta_{W1}) \qquad \text{(II-5)}$$

where: H_z = direct irradiance normal to beam at depth z
z = see figure II-2 (p. 20)

Effects due to volume scattering can be developed by considering the behavior of the radiation in a differential volume, dV. The radiation returned to a point on the surface from the incremental volume depends in part on the angle between the light beam and the ray from the incremental volume to the point on the surface. This dependence is specified by the scattering function, $\beta(\alpha)$,

given by the definition

$$\beta(\alpha) \equiv \frac{1}{H_z} \frac{dJ}{dV} \qquad \text{(II-6a)}$$

$$dJ = H_z \beta(\alpha) dV \qquad \text{(II-6b)}$$

where: $\beta(\alpha)$ = scattering function

α = viewing angle

dJ = scattered light intensity

dV = incremental volume

Considering attenuation and dispersion of the signal, the incremental irradiance at the surface due to radiation scattered from an incremental volume is

$$dH_a = dJ \frac{\exp(-c_A R)}{R^2} \qquad \text{(II-7)}$$

where: dH_a = irradiance at point below surface due to dV

R = distance from dV to point below surface

Although multiple scattering occurs, light that has been scattered for the second time is three orders of magnitude less than that scattered once and, hence, may be neglected in most instances.

To see the effects of waste concentration on scattered radiance, (II-7) must be expressed in terms of simpler

variables. Substituting (II-6b) for dJ gives

$$dH_a = H_z \beta(\alpha) dV \frac{\exp(-c_A R)}{R^2} \qquad \text{(II-8)}$$

To expand this, simple expressions for H_z, $\beta(\alpha)$ and dV are needed. Combining (II-3) with (II-5) gives an expression for H_z. The expression is

$$H_z = H_s \frac{1-\rho_a}{\cos\theta_{w1}} \exp(-c_A z \sec\theta_{w1}) \qquad \text{(II-9)}$$

For $\beta(\alpha)$, figure II-3 suggests the approximation

$$\beta(\alpha) = \beta_o(\alpha)(1 + K'W) \qquad \text{(II-10)}$$

where: $\beta_o(\alpha)$ = scattering function for zero contaminant concentration

K' = constant for each contaminant

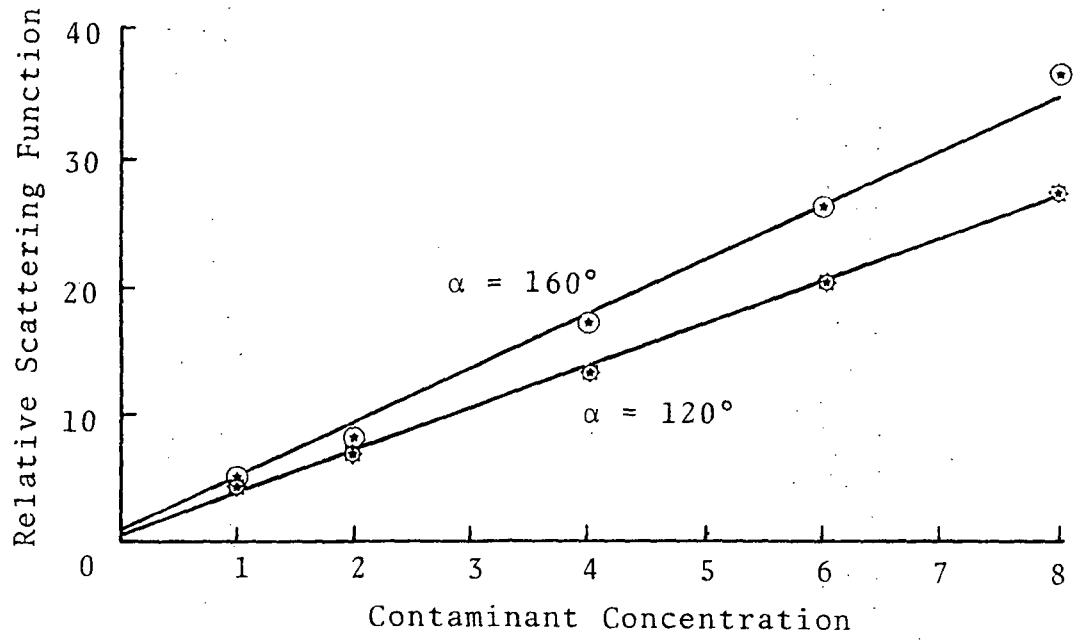
The incremental volume, dV, may be expressed as

$$dV = R^2 dR\, d\Omega \qquad \text{(II-11)}$$

where: $d\Omega$ = solid angle at point on surface subtended by dV

Two expressions that will be needed are

$$R = z \sec\theta_{w2} \qquad \text{(II-12)}$$



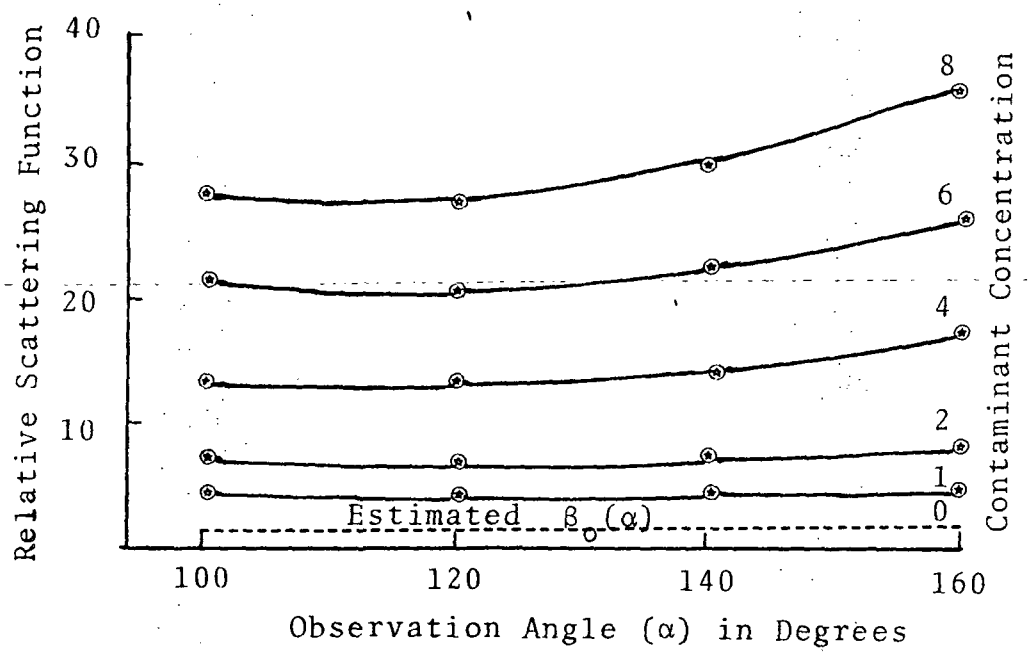


Fig. II-3 Scattering Function Curves [10]

where: θ_{W2} = see figure II-2 (p. 20)

and

$$dR = \frac{dz}{\cos \theta_{W2}} \qquad \text{(II-13)}$$

Substituting (II-9), (II-10) and (II-11) for, respectively, H_z, $\beta(\alpha)$ and dV in (II-8) gives

$$dH_a = \left(H_s \frac{1-p_a}{\cos \theta_{W1}} \exp\left(-c_A z \sec \theta_{W1}\right) \right) \qquad \text{(II-14)}$$

$$\times \left(\beta_o(a)(1+K'W)\, dR\, d\Omega \exp(-c_A R)\right)$$

Substituting (II-4), (II-12) and (II-13) for, respectively, c_A, R and dR, then rearranging gives

$$dH_a = \left(H_s \frac{\beta_o(a)(1-p_a)\, d\Omega}{\cos\theta_{W1} \cos\theta_{W2}} \right) \qquad \text{(II-15)}$$

$$\times \left((1+K'W) \exp(-(a+bW)(\sec\theta_{W1} + \sec\theta_{W2})z)\, dz\right)$$

The irradiance at a point just below the surface of the water due to scattering from all incremental volumes at view angle α is

$$H_a = \left(H_s \frac{\beta_o(a)(1-p_a)\, d\Omega}{\cos\theta_{W1} \cos\theta_{W2}} \right) \qquad \text{(II-16)}$$

$$\times \left(\int (1+K'W) \exp(-(a+bW)(\sec\theta_{W1} + \sec\theta_{W2})z)\, dz\right)$$

Under certain conditions, (II-16) may be evaluated. If the variation in W with z is known, (II-16) may be integrated numerically. If W is independent of z, that is, if W is a constant, the integration may be performed directly. If the exponential approaches zero as z approaches the maximum depth contributing scattered irradiance, direct integration and simplification gives

$$H_a = H_s \frac{\beta_o(\alpha)(1-p_a)(1+K'W)\,d\Omega}{(a+bW)(\cos\theta_{W1}+\cos\theta_{W2})} \qquad \text{(II-17)}$$

The radiance associate with H_a is

$$N_W = \frac{H_a}{d\Omega} \qquad \text{(II-18a)}$$

$$N_W = H_s \frac{\beta_o(\alpha)(1-p_a)(1+K'W)}{(a+bW)(\cos\theta_{W1}+\cos\theta_{W2})} \qquad \text{(II-18b)}$$

where: N_W = radiance at point just below the surface of the water due to scattering from all incremental volumes at view angle α

As the radiation passes from the water to the atmosphere, part is reflected and the remainder is spread through a larger angle to produce a radiance

$$N_a = N_W \frac{(1-p_o)}{n^2} \qquad \text{(II-19a)}$$

$$N_a = H_s \frac{\beta_o(\alpha)(1-p_a)(1-p_w)(1+K'W)}{n^2(a+bW)(\cos\theta_{W1}+\cos\theta_{W2})} \qquad \text{(II-19b)}$$

where: N_a = radiance at point just above the surface of the water due to N_w

p_w = fraction of radiation reflected at water-air interface

Equation (II-19b) shows the relation between scattered radiance just above the surface of the water and contaminants in the water.

Sources of Noise and Interference. The noise and interference associated with scattered radiation have several sources. The first source is the atmosphere. As radiation passes through the atmosphere, part of it is scattered and part of it is absorbed. The irradiance remaining when the radiation reaches the surface of the water is given by

$$H_s = H_o \cos\theta_{AI} \exp(-A \sec\theta_{AI}) \qquad \text{(II-20)}$$

where: H_o = irradiance at the surface of the atmosphere

A = extinction optical thickness

The second source of interference is multiple scatter in the water. As mentioned previously, the magnitude of radiation that is scattered for the second time is three orders of magnitude less than that scattered once, therefore, it may be neglected.

The third source of interference is specular reflection of sunlight at the surface of the water. Because

this reflection is partially polarized, interference may be reduced by using polarizing filters. The reflection can be avoided by using an oblique view for the sensor or by proper selection of the zenith angle. As shown in figure II-4, the zenith angle should be selected to obtain

$$\theta_{min} = \phi_{I_{max}} + 2\delta_{max} \qquad \text{(II-21)}$$

where: θ_{min} = minimum acceptable sun zenith angle

$\phi_{I\,max}$ = sensor viewing angle

δ_{max} = maximum slope of water surface

With this sun zenith angle, all specular reflection will be at angles greater than the maximum angle in the field of view of the sensor.

The fourth source of interference is skylight, the sunlight scattered by particles in the atmosphere. Some skylight is scattered directly to the sensor, and some is diffusely reflected to the sensor from the water. The scattering is proportional to $1/\lambda^4$, where λ is the wavelength of the radiation, for small particles and becomes less selective as the size of particles in the atmosphere increases. This results in the scattered radiation being predominantly blue and its effects may be reduced by a minus blue filter. As the particle size increases, longer

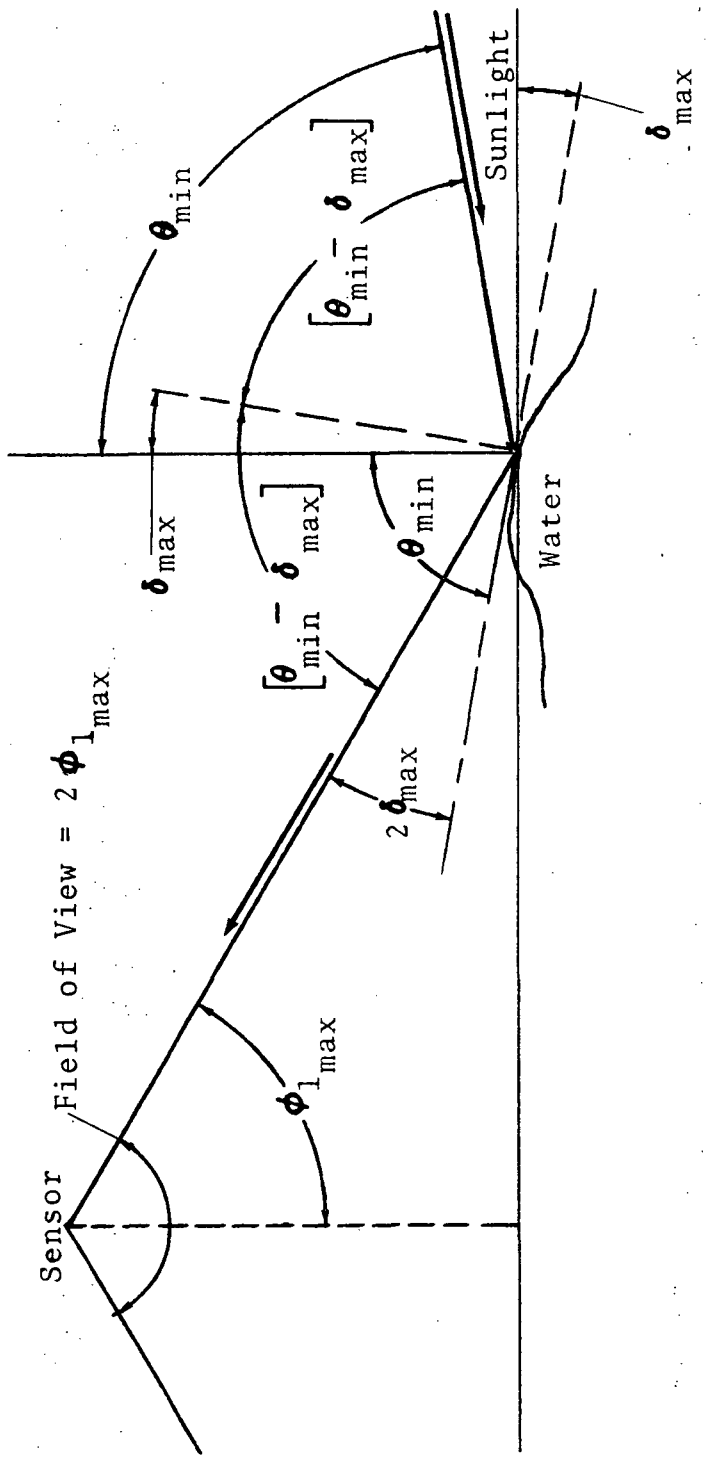


Fig. II-4 Geometry for Selection of Minimum Zenith Angle

wavelengths may also need to be filtered.

A fifth source of interference is path radiance of the atmosphere. This increases with altitude and can be half the measured radiation at 10,000 feet. Unfortunately, no method is currently available to take this effect into account so it must be neglected.

The sixth source of interference is reflection from the bottom of the water body. If the water is sufficiently deep, this effect may be ignored.

The equations discussed previously can be modified to account for atmospheric scattering and absorption. Substituting (II-20) for H_s in (II-19b) gives

$$N_a = H_o \frac{\beta_o(\alpha)(1-p_o)(1-p_w)\cos\theta_{A1}}{n^2(\cos\theta_{W1}+\cos\theta_{W2})} \frac{1+K'W}{a+bW} \exp(-A\sec\theta_{A1}) \quad \text{(II-22)}$$

Representing the factors that are independent of atmosphere or wastes by a single factor gives

$$N_a = Y \frac{1+K'W}{a+bW} \exp(-A\sec\theta_{A1}) \quad \text{(II-23)}$$

where: $$Y = H_o \frac{\beta_o(\alpha)(1-p_o)(1-p_w)\cos\theta_{A1}}{n^2(\cos\theta_{W1}+\cos\theta_{W2})}$$

Photography

Photography is essentially a chemical change produced by exposure to radiation. A lens system is used to display an image on a film coated with an emulsion. The radiation produces a chemical reaction in the emulsion that increases with the image intensity and the length of time the image is displayed. During development, the unreacted portion of the emulsion is removed. The opacity or density of the remaining portion depends on the amount of reacted emulsion [11].

Spectral signatures are recorded by photographic means by the use of different film-filter combinations. Films coated with different emulsions respond to different bands of radiation. For color photography, three emulsion layers are placed on one film. Each emulsion responds to a different band of light and produces a different color when it is developed. A "signature" is obtained from the density of each emulsion [11].

Filters, which allow only certain bands of radiation to pass, usually respond to much narrower, more precise bands than those to which emulsions respond. They are used for multi-band photography; several photographs of the same scene taken simultaneously with different film-filter combinations. A "signature" is obtained from the

densities of the emulsions in the different photographs [11].

Radiation Reaching Film

In order to relate photographic data to waste concentrations, the relation between scattered light and photographic data must be known. This relation will be developed by relating the irradiance on the photograph to the radiance on the surface, then by developing the response of the film to the radiance.

The ratio between an incremental area on the ground, dS, and the corresponding incremental area in a photograph, dS', partially determines the irradiance on the photograph. The relationship is represented in figure II-5. Expressions for dS and dS' are

$$dS = \frac{\left[\frac{z_o}{\cos \phi_1}\right]^2 d\xi}{\cos \phi_1} \qquad \text{(II-24)}$$

where: dS = incremental area on ground

z_o = altitude of airplane

$d\xi$ = solid angle at P (see figure II-5(p.34)) substended by dS

ϕ_1 = see figure II-5 (p. 34)

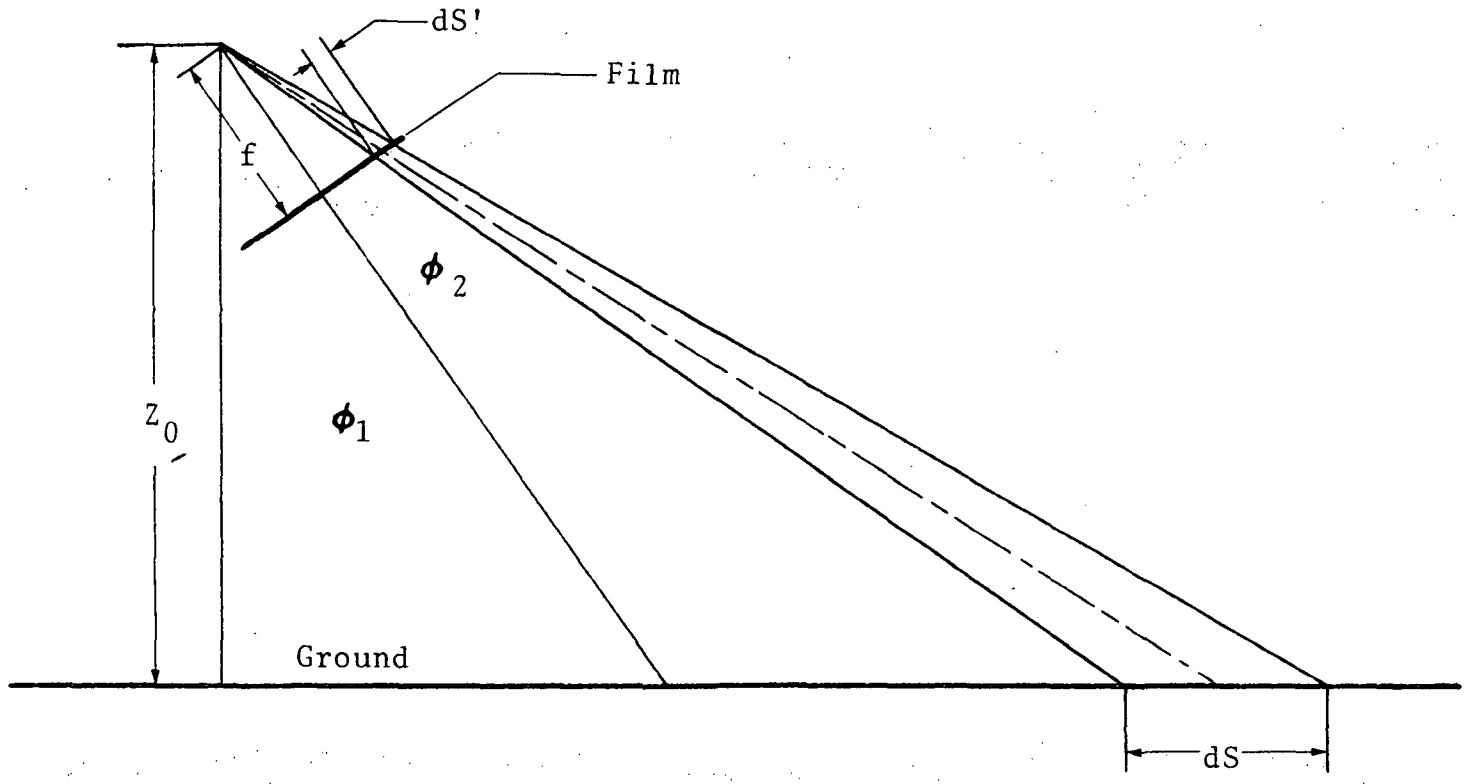


Fig. II-5 Representation of Relation Between Incremental Film Area and Incremental Ground Area

and

$$dS' = \frac{\left[\frac{f}{\cos \phi_2}\right]^2 d\xi}{\cos \phi_2} \qquad \text{(II-25)}$$

where: dS' = incremental area in photograph corresponding to dS

f = focal length of camera (see figure II-5 (p. 34))

ϕ_2 = see figure II-5 (p. 34)

The ratio of (II-24) and (II-25) gives the desired expression [10]

$$\frac{dS}{dS'} = \left[\frac{Z_o}{f}\right]^2 \left[\frac{\cos \phi_2}{\cos \phi_1}\right]^3 \qquad \text{(II-26)}$$

Another factor that influences the irradiance on the photography is the solid angle at the ground subtended by the lens of the camera. From dS, a lens of diameter D would appear as an ellipse with major axis D/2 and minor axis D $\cos\phi_2/2$ (refer to figure II-6). The area of the ellipse would be

$$Q = \frac{\pi}{2} D^2 \cos \phi_2 \qquad \text{(II-27)}$$

The distance from dS to the lens is $Z_o/\cos\phi_1$, hence, the solid angle at dS subtended by the lens is

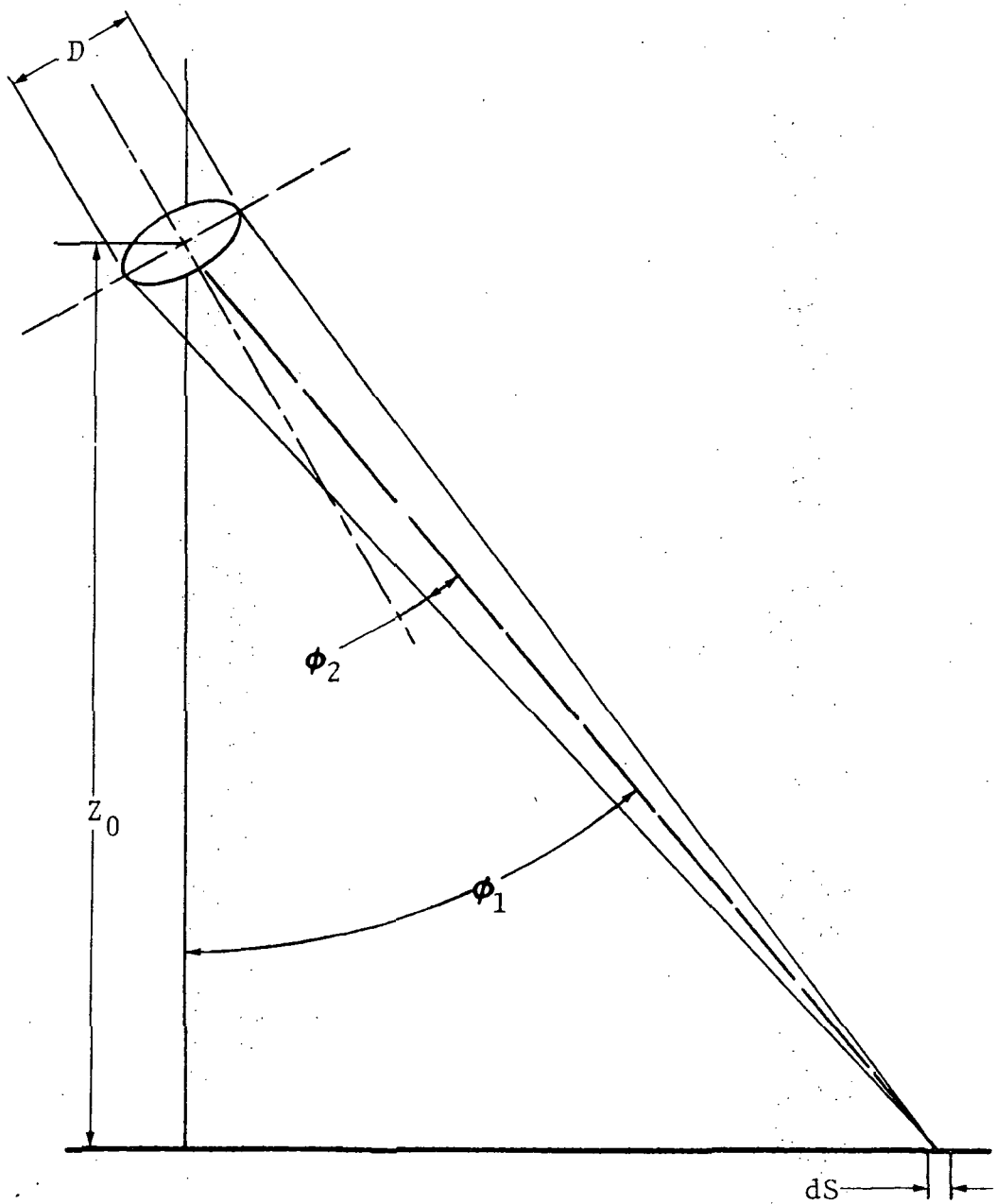


Fig. II-6 Solid Angle Subtended by Lens

$$d\psi = \frac{\pi}{4}\left[\frac{D^2 \cos^2\phi_1}{Z_o}\right] \cos\phi_2 \qquad \text{(II-28)}$$

where: $d\psi$ = solid angle at dS subtended by lens

The radiant flux at the lens due to scattered radiance at dS is

$$dp' = N_a \exp(-A' \sec\phi_1)\, dS \cos\phi_1\, d\psi \qquad \text{(II-29)}$$

where: dp' = radiant flux at lens due to dS

A' = extinction optical thickness from water surface to lens

The irradiance on the film is

$$H' = T_{Lens}\, \frac{dp'}{dS'} \qquad \text{(II-30a)}$$

$$H' = T_{Lens} N_a \exp(-A' \sec\phi_1) \cos\phi_1\, d\psi\, \frac{dS}{dS'} \qquad \text{(II-30b)}$$

where: H' = irradiance on the film

T_{Lens} = transmission factor for lens

Substituting (II-26) and (II-28) for dS/dS' and $d\psi$, respectively, gives

$$H' = N_a \exp(-A' \sec\phi_1)\left(\frac{\pi}{4}\,\frac{T_{Lens}}{(f/D)^2}\right)\cos^4\phi_2 \qquad \text{(II-31)}$$

This may be written as

$$H' = Y' N_a \exp(-A' \sec \phi_1) \cos^4 \phi_2 \qquad \text{(II-32a)}$$

$$\frac{H'}{N_a} = Y' \exp(-A' \sec \phi_1) \cos^4 \phi_2 \qquad \text{(II-32b)}$$

where: $Y' = \frac{\pi}{4} \frac{T_{Lens}}{(f/D)^2}$

The relationship between the irradiance on the film and the radiance immediately above the surface of the water is given by (II-32a) [10].

Response of Film to Radiation

The response of film to radiation depends on exposure and the characteristic curve of the film. If extreme values of exposure time are avoided, the extent of the reaction of the emulsion on a film depends on its exposure,

$$E = H' \times T \qquad \text{(II-33)}$$

where: E = exposure

T = duration of irradiance

Obviously, any exposure may result from any one of an infinite number of combinations of irradiance and exposure

time [10].

The extent of the reaction of an emulsion is specified by its density, which is defined as

$$D_o = \log_{10}\left[\frac{I_o}{I}\right] \qquad \text{(II-34)}$$

where: D_o = density
I_o = incident radiation
I = transmitted radiation

Density and exposure are related by a curve (figure II-7) that is characteristic of the film. The straight line portion of the curve may be represented by

$$D_o = M + G \ln(E) \qquad \text{(II-35)}$$

where: M = intercept of straight line with abscissa
G = slope of straight line portion [10]

From (II-35), an expression for E is

$$E = \exp\left(\frac{D - M}{G}\right) \qquad \text{(II-36)}$$

Substituting (II-35) for E and dividing by T gives

$$H' = \frac{\exp((D_o - M)/G)}{T} \qquad \text{(II-37)}$$

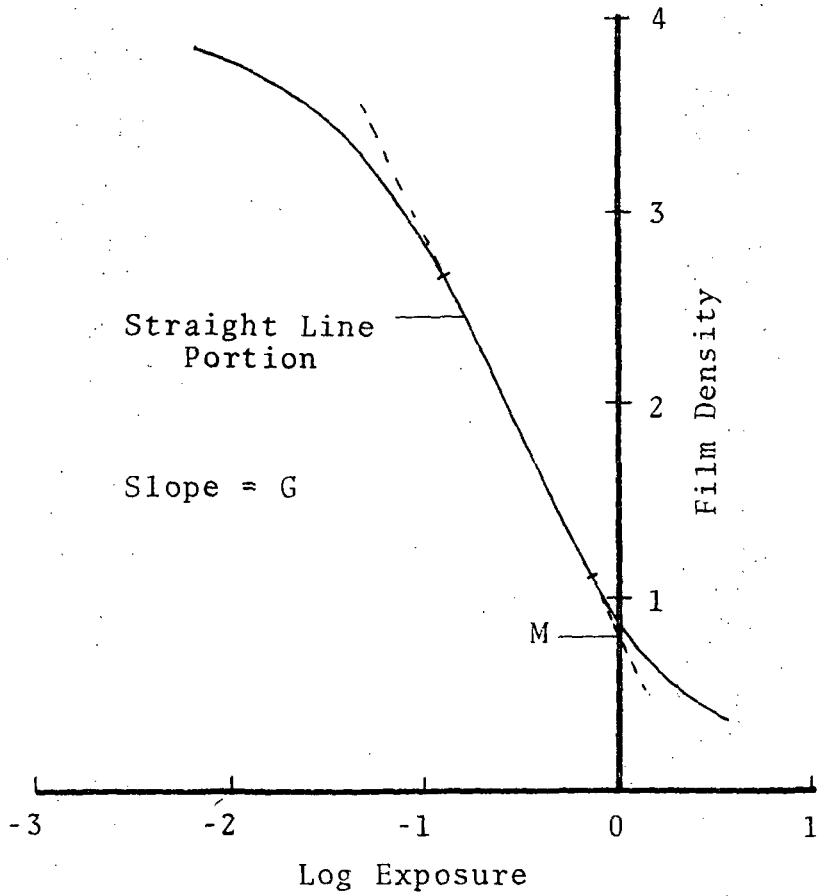


Fig. II-7 Typical Characteristic Curve of Film

Equation (II-37) shows the response of the film to radiation by relating the irradiance on the film to the density of the developed photograph.

Dividing (II-37) by (II-32b) gives

$$N_a = \frac{\exp((D_o - M)/G)\exp(A' \sec\theta_{A2})}{Y' T \cos^4\phi_2} \qquad \text{(II-38)}$$

an expression relating scattered radiance at the surface of the water to film density [10].

Multispectral Scanners

A multispectral scanner is illustrated in figure II-8. The heart of the scanner is a mirror mounted as a 45 degree angle with a shaft. As the shaft, mounted parallel to the velocity vector, rotates, the mirror scans perpendicular to the flight line. The mirror reflects light from the ground through a mirror system to a spectrometer [12,13, 14].

Light enters the spectrometer through an aperture and is spread into its spectrum by a prism as shown in figure II-8 (p. 42). The radiance of any given band may be obtained by placing a suitable sensor in the position of the spectrum corresponding to that band. Two advantages of this system are: one, it can be used to measure radiances

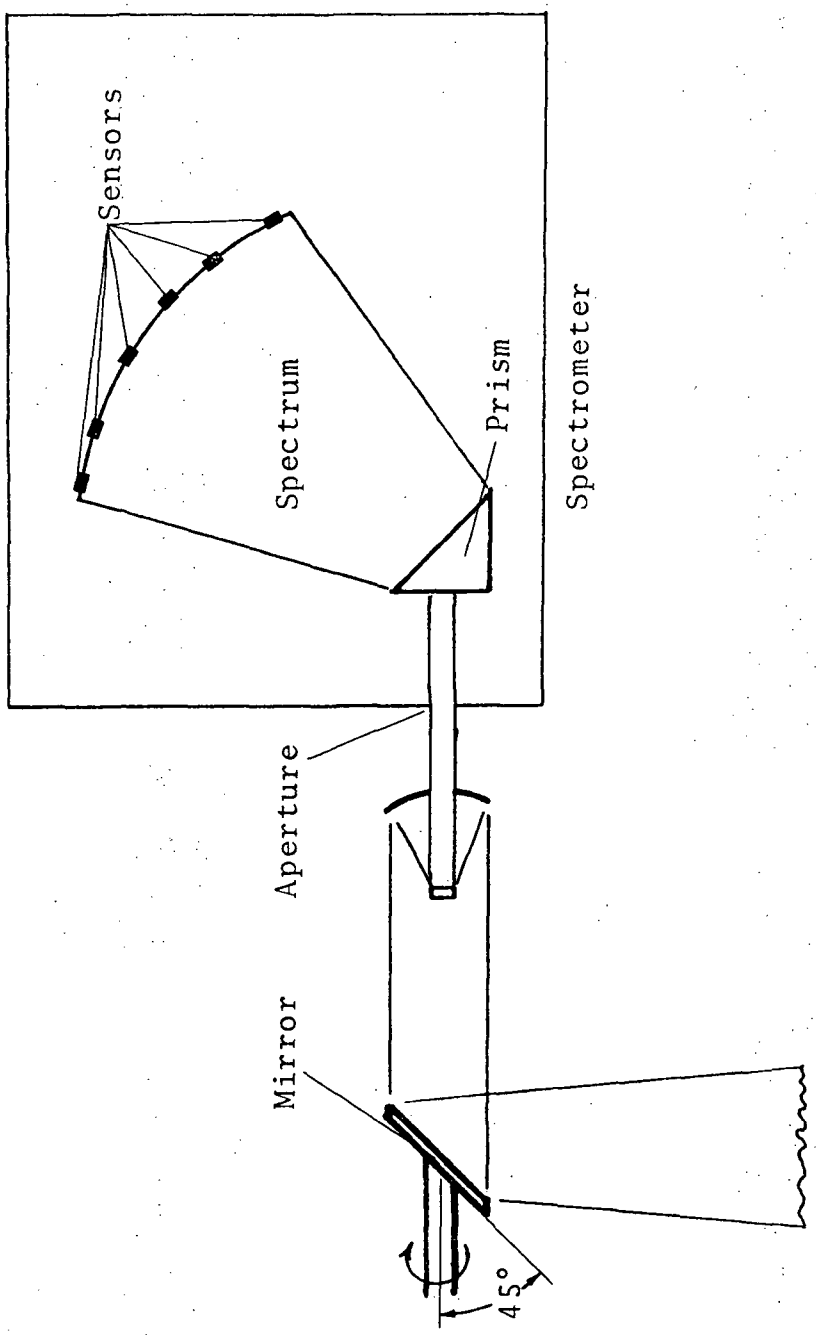


Fig. II-8 Representation of Multispectral Scanner

in more bands simultaneously, and two, diverse sensors may be used to measure radiances in different portions of the spectrum. The sensors produce electrical signals proportional to the radiance of their bands. These signals are then recorded for later processing [7,12,13,14].

A scanner may be calibrated during the portion of each scan that its mirror faces the scanner housing. Signals corresponding to known radiances may be obtained by inserting sources such as lamps and skylights into the housing. These signals provide a range of known radiances for calibration [12].

CHAPTER III

DATA ANALYSIS TECHNIQUES

The purpose of regression analysis is to use empirical data to obtain a mathematical expression relating a dependent variable to one or more independent variables. The first step in performing a regression is to assume a model, i.e. a form for the equation. The second step is to use empirical data to estimate the model parameters or the unknowns in the equation. The third step is to test the reliability of the model. The use of regression analysis is described by Draper and Smith [16], and except where noted, the following developments are based on that work.

Regression Model

Model Selection

A regression model is obtained from theoretical considerations, an assumption, or scatter diagrams. Theoretical considerations may provide an equation relating dependent and independent variables, and that equation can serve as the model. Models based on experience or a simple relationship that seems reasonable are often obtained by assumption. When scatter diagrams are used, the dependent variable is plotted versus each independent

variable. The relationship between the dependent and independent variables is then estimated from the plots.

A model often used in regression analysis is the linear model. This model, which has the form

$$y = \beta_0 + \sum_{j=1}^{p} x_j \beta_j + \epsilon \qquad \text{(III-1)}$$

where: y = dependent variable

x_j = jth independent variable

p = number of independent variables

β_0 = constant term

β_j = constant coefficient of jth independent variable (β_0 and the β_j's are the "parameters" of the model)

This assumes that the value of the dependent variable varies in direct proportion to the value of each of the independent variables. This model is advantageous because methods of estimating β_0 and the β_j's and methods of testing the model are available, and the model is easily understood and used.

The linear model is widely applicable because many relationships are approximately linear for some range of values and many nonlinear relationships can be linearized. Nonlinear relationships between a dependent and independent variable are linearized by generating one or more new independent variable. The new variables, obtained as

functions of the original independent variable, are used in the model. The functions used to generate the new variables are determined more by "art" than by "science", but scatter diagrams are often useful.

Matrix notation is convenient when discussing regression. In this discussion, the notation $\underset{\sim}{A}$ represents a matrix, $\underset{\sim}{A}^T$ represents the transpose of $\underset{\sim}{A}$, and $\underset{\sim}{A}^{-1}$ represents the inverse of $\underset{\sim}{A}$. Using this notation, the linear regression model is

$$\underset{\sim}{Y} = \underset{\sim}{X}\underset{\sim}{\beta} + \underset{\sim}{\epsilon} \qquad \text{(III-2)}$$

where: $\underset{\sim}{Y} = [y_1 \cdots y_n]^T$

$$\underset{\sim}{X} = \begin{bmatrix} x_{10} & \cdots & x_{1p} \\ \vdots & & \vdots \\ x_{n0} & \cdots & x_{np} \end{bmatrix}$$

$\underset{\sim}{\beta} = [\beta_0 \cdots \beta_p]^T$

$\underset{\sim}{\epsilon} = [\epsilon_1 \cdots \epsilon_n]^T$

y_i = observed value of dependent variable for ith data point

$x_{i0} = 1$

x_{ij} = observed value of jth independent variable for ith data point

β_j = constant coefficient of jth term

ϵ_i = error term for ith data point

n = number of data points ($n \geq p+1$)

p = number of independent variables

Parameter Calculation

The parameters of (III-2) are represented by $\underset{\sim}{\beta}$ and are approximated by $\underset{\sim}{B}$. It is unreasonable to expect to calculate the exact value of any dependent variable from a set of independent variables and the model accounts for this by including the random error term, $\underset{\sim}{\epsilon}$. To minimize the error, a least squared error criterion for selecting $\underset{\sim}{B}$ is used. This calculation assumes that the value of $\underset{\sim}{Y}$ independent of error is

$$\underset{\sim}{\hat{Y}} = \underset{\sim}{X}\,\underset{\sim}{B} \qquad \text{(III-3)}$$

where: $\underset{\sim}{\hat{Y}} = [\hat{y}_1 \cdots \hat{y}_n]^T$

$\underset{\sim}{B} = [b_0 \cdots b_p]^T$

$\hat{y}_i$ = calculated value of dependent variable for ith data point

b_j = estimated value of β_j

From (III-2) and (III-3), the error is

$$\underset{\sim}{\epsilon} = \underset{\sim}{Y} - \underset{\sim}{\hat{Y}} \qquad \text{(III-4a)}$$

$$\underset{\sim}{\epsilon} = \underset{\sim}{Y} - \underset{\sim}{X}\,\underset{\sim}{B} \qquad \text{(III-4b)}$$

The sum of the squared error is

$$SSE = \underset{\sim}{\epsilon}^T \underset{\sim}{\epsilon} \qquad \text{(III-5a)}$$

$$SSE = (\underset{\sim}{Y} - \underset{\sim}{X}\,\underset{\sim}{B})^T(\underset{\sim}{Y} - \underset{\sim}{X}\,\underset{\sim}{B}) \qquad \text{(III-5b)}$$

where: SSE = sum of the squared error

The $\underset{\sim}{B}$ that minimizes SSE is the one that satisfies the normal equation,

$$(\underset{\sim}{X}^T\underset{\sim}{X})\underset{\sim}{B} = \underset{\sim}{X}^T\underset{\sim}{Y} \quad , \tag{III-6}$$

obtained by taking the partials of (III-5b) with respect to each b_j, setting the partials to zero, and representing the resulting system of equations in the form of matrices. The solution of (III-6) for $\underset{\sim}{B}$ is

$$\underset{\sim}{B} = (\underset{\sim}{X}^T\underset{\sim}{X})^{-1}\underset{\sim}{X}^T\underset{\sim}{Y} \tag{III-7}$$

Analysis of Variance

Once the parameters of a regression are calculated, the question of whether or not the regression is significant arises. One way to obtain some indication of the significance of the regression is to perform an analysis of variance which shows how much of the variation in the dependent variable can be predicted by the regression.

The variation in the dependent variable is measured by the sum of squares total given by

$$SST = \underset{\sim}{Y}^T\underset{\sim}{Y} \tag{III-8}$$

where: SST = sum of squares total

The simplest and most direct method of estimating $\underset{\sim}{Y}$ would be to use the mean value of y_i given by

$$\bar{y} = \sum_{i=1}^{n} y_i / n \qquad \text{(III-9)}$$

where: $\bar{y}$ = mean value of y_i

The portion of the sum of squares for which this method would account is

$$SSM = n\bar{y}^2 \qquad \text{(III-10)}$$

where: SSM = sum of squares due to mean

The remaining portion of the sum of squares is

$$SST_M = SST - SSM \qquad \text{(III-11a)}$$

$$SST_M = \underset{\sim}{Y}^T\underset{\sim}{Y} - n\bar{y}^2 \qquad \text{(III-11b)}$$

where: SST_M = sum of squares total adjusted for mean

It is this sum of squares for which a regression attempts to account

The sum of squares for which a regression accounts is

$$SSR = \underset{\sim}{B}^T\underset{\sim}{X}^T\underset{\sim}{Y} \qquad \text{(III-12)}$$

where: SSR = sum of squares due to regression

Only the portion of SSR that is greater than SSM justifies the regression, hence, the important value is

$$SSR_M = SSR - SSM \qquad \text{(III-13a)}$$

$$SSR_M = \underset{\sim}{B}^T \underset{\sim}{X}^T \underset{\sim}{Y} - n\bar{y}^2 \qquad \text{(III-13b)}$$

where: SSR_M= sum of squares due to regression adjusted for mean

The portion of SST_M for which SSR_M does not account is the sum squared error. Because the total sum of squares must equal the sum of the sums of squares contributing to it, the sum of squared error is

$$SSE = SST - SSR \qquad \text{(III-14a)}$$

$$SSE = SST_M - SSR_M \qquad \text{(III-14b)}$$

$$SSE = \underset{\sim}{Y}^T \underset{\sim}{Y} - \underset{\sim}{B}^T \underset{\sim}{X}^T \underset{\sim}{Y} \qquad \text{(III-14c)}$$

This is a more convenient computational form than (III-5b).

Two more values of interest to an analysis of variance are mean square of the regression and mean square of the error. To calculate these values, the degrees of freedom of SSR_M and SSE are needed.

Degrees of freedom indicate how many variables are required to calculate a value. For SST, SSM and SSR, the degrees of freedom are, respectively, n, 1, and (p+1).

Because the degrees of freedom of a sum of sums of squares must equal the sum of the degrees of freedom of the sums of squares,

$$DF_{RM} = DF_R - DF_M = p \qquad \text{(III-15a)}$$

$$DF_{TM} = DF_T - DF_M = n - 1 \qquad \text{(III-15b)}$$

$$DF_E = DF_T - DF_R = n - p - 1 \qquad \text{(III-15c)}$$

$$DF_E = DF_{TM} - DF_{RM} = n - p - 1 \qquad \text{(III-15d)}$$

where: DF_{RM} = degrees of freedom of SSR_M

DF_R = degrees of freedom of SSR

DF_M = degrees of freedom of SSM

DF_{TM} = degrees of freedom of SST_M

DF_T = degrees of freedom of SST

DF_E = degrees of freedom of SSE

Using (III-15a) and (III-15d), the mean square of the regression is

$$MSR_M = SSR_M / DF_{RM} \qquad \text{(III-16)}$$

where: MSR_M = mean square of the regression corrected for the mean

and the mean square of the error is

$$MSE = SSE / DF_E \qquad \text{(III-17)}$$

where: MSE = mean square of the error

These values indicate, respectively, the average portion

of the sum of squares attributable to each degree of freedom of the regression and to each degree of freedom of the error.

The most convenient method of arranging the data for an analysis of variance is an analysis of variance table as illustrated by figure III-1.

Analysis of Variance

Source	Sum of Squares	Degrees of Freedom	Mean Square
Mean	SSM	DF_M	
Regression Adjusted for Mean	SSR_M	DF_{RM}	MSR_M
Error	SSE	DF_E	MSE
Total Adjusted for Mean	SST_M	DF_{TM}	

Fig. III-1 *One Arrangement for an Analysis of Variance Table*

Testing Regression Model

Normal Error

The analysis of variance discussed above provides little indication of the usefulness of the regression, it does little more than indicate whether or not the regression accounts for more of the sum of squares than the mean alone. Further tests require assumptions about the distribution of the error. A common assumption is that the error is normally distributed with mean of zero and variance of σ^2, i.e., the probability density function for each ϵ_i is

$$f(\epsilon_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{\epsilon_i^2}{2\sigma^2}\right] \qquad \text{(III-18)}$$

where: $f(\epsilon_i)$ = probability density function of ϵ_i

σ = standard deviation of ϵ_i

The interpretation of this is that if the data for the ith data point were collected a large number of times, the average value for ϵ_i would approach zero and the average value of ϵ_i^2 would approach σ^2.

Normal error is commonly assumed for three reasons. First, errors in many natural processes are normal. Second, the errors are often due to an accumulation of other errors, and the Central Limit Theorem states that such cummulative errors will tend to be normal, regardless of the distributions of the contributing errors. Third, the tests

requiring a normal error distribution are "robust", i.e. even if the error is not quite normal, only small discrepencies are produced in the tests.

If normal error is assumed, it justifies the method used to obtain $\underset{\sim}{B}$. If the error is normal, the probability density function for $\underset{\sim}{Y}$ is

$$f = \frac{\exp[-\underset{\sim}{\epsilon}^T\underset{\sim}{\epsilon}/2\sigma^2]}{\sigma^n\ (2\pi)^{n/2}} \qquad \text{(III-19)}$$

where: = probability density function for Y

The value of f is maximized by minimizing $\underset{\sim}{\epsilon}^T\underset{\sim}{\epsilon}$. Because the method used to determine $\underset{\sim}{B}$ minimized $\underset{\sim}{\epsilon}^T\underset{\sim}{\epsilon}$, it maximized the probability density function for $\underset{\sim}{Y}$.

Distributions

The regression tests described in the following discussion will make use of three probability distributions: the normal distribution, the t-distribution, and the F-distribution as shown in figure III-2. Such distributions are used to determine the probability that certain random variables lie within given limits. If F(z) is the distribution function for a random variable z, it has the properties: (1) $F(-\infty) = 0$ and $F(\infty) = 1$, (2) if $z_1 < z_2$ then $F(z_1) \leq F(z_2)$, and (3) F(z) is continuous from the right. If $z_1 < z_2$, then the probability that $z_1 \leq z \leq z_2$ is

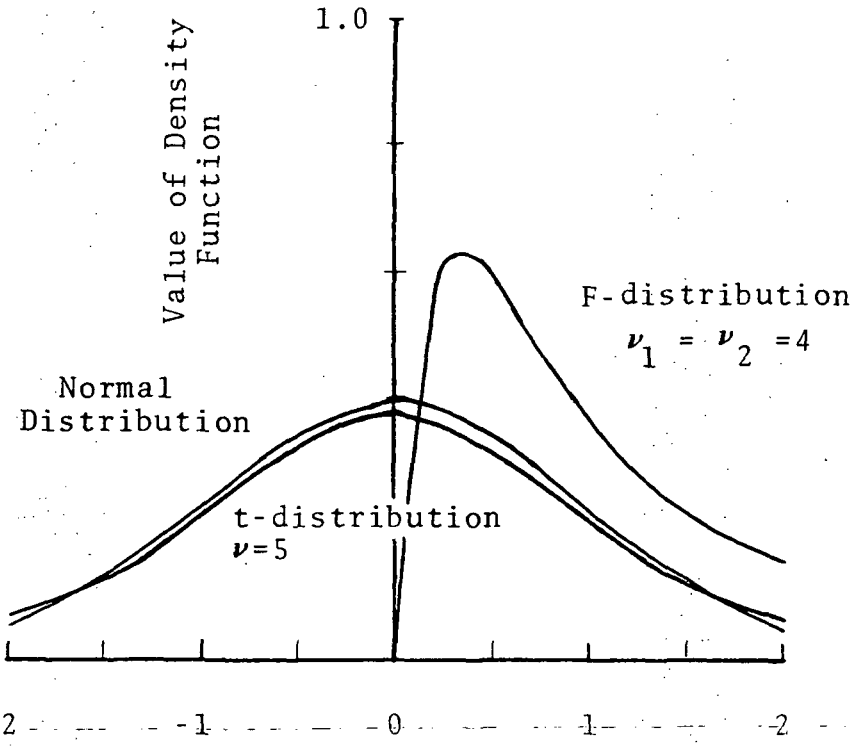


Fig. III-2 Distribution Functions

$(F(z_2) - F(z_1))$ [20].

Values for common distribution functions are usually obtained from tables such as those in Standard Mathematical Tables [21]. Tabulated data for normal distributions and t-distributions are for standardized distributions described later. A normal or t-distribution can be standardized by

$$y = \frac{x - \mu}{\sigma} \qquad \text{(III-20)}$$

where: y = random variable with standarized normal distribution (standarized t-distribution)

x = random variable with normal distribution (t-distribution)

μ = mean value of x

σ = standard deviation of x

The standarized normal distribution is given by [21]

$$F(z) = \int_{-\infty}^{z} \left[\frac{\exp(-t^2)}{2\pi}\right]^{1/2} dt \qquad \text{(III-21)}$$

where: $F(z)$ = distribution function of z

z = maximum value of the random variable in the interval

t = dummy variable of integration

Because the distribution is symmetrical about $z = 0$, $F(0) = .5$. Often, tables only give values for $0 \leq z$, but values for $z < 0$ can be obtained from [21]

$$F(z) = 1 - F(-z) \tag{III-22}$$

The standarized t-distribution is given by [21]

$$F(t(\nu)) = \int_{-\infty}^{t(\nu)} \frac{\Gamma((\nu+1)/2)}{(n\pi)^{1/2}\Gamma(\nu/2)} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}} dx \tag{III-23}$$

where: $F(t(\nu))$ = distribution function of $t(\nu)$

$t(\nu)$ = maximum value of the random variable in the interval

ν = degrees of freedom of $t(\nu)$

x = dummy variable of integration

As in the case of the normal distribution, $F(t(\nu)) = .5$ if $t(\nu) = 0$, and (III-22) can be used to evaluate $F(t(\nu))$ if $t(\nu) < 0$.

Unlike the normal distribution, the value of $F(t(\nu))$ depends on ν. Tabulated data usually gives values of $t(\nu)$ corresponding to various values of $F(t(\nu))$ and ν. Most of the values for ν are for $\nu \leq 30$. For $30 < \nu$, the t-distribution can be approximated by a normal distribution [17].

The F-distribution applies to the ratio of two chi-square random variables, each divided by its degrees of freedom, as in (III-24).

$$F(\nu_1,\nu_2) = \frac{x_1/\nu_1}{x_2/\nu_2} \qquad \text{(III-24)}$$

where: $F(\nu_1,\nu_2)$ = random variable with an F-distribution

x_1 = random variable with a chi-square distribution

x_2 = random variable with a chi-square distribution

ν_1 = degrees of freedom of x_1

ν_2 = degrees of freedom of x_2

For the purpose of this discussion, an explanation of a chi-square variable is deemed unnecessary except to say that SSR_M and SSE are chi-square variables with, respectively, p and n-p-1 degrees of freedom [17].

The F-distribution is given by [21]

$$F(F(\nu_1,\nu_2)) = \int_0^{F(\nu_1,\nu_2)} \frac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)} \nu_1^{\frac{\nu_1}{2}} \nu_2^{\frac{\nu_2}{2}} x^{\frac{\nu_1-2}{2}} (\nu_2+\nu_1 x)^{-\frac{\nu_1+\nu_2}{2}} dx$$

(III-25)

where: $F(F(\nu_1,\nu_2))$ = distribution function of $F(\nu_1,\nu_2)$

$F(\nu_1,\nu_2)$ = maximum value of the random variable in the interval

x = dummy variable of integration

Tables usually give values of $F(\nu_1,\nu_2)$ corresponding to particular values of $F(F(\nu_1,\nu_2))$, ν_1, and ν_2.

Application of Tests

One test of a regression is the correlation coefficient given by

$$R = \left(\frac{SSR_M}{SST_M}\right)^{1/2} \qquad \text{(III-26)}$$

where: R = correlation coefficient

The value of R^2 is the fraction of SST_M explained by SSR_M, and it indicates the strength of the relation between the dependent variable and the regression [17].

One test of significance of a regression can be based on the F-distribution because SSR_M with degrees of freedom p and SSE with degrees of freedom n-p-1 are chi-square random variables and satisfy the conditions for an F-distribution of the form

$$F(p,n-p-1) = \frac{SSR_M / p}{SSE / (n-p-1)} \qquad \text{(III-27a)}$$

$$F(p,n-p-1) = \frac{MSR_M}{MSE} \qquad \text{(III-27b)}$$

A regression can usually be accepted as a useful predictor of a dependent variable with α probability if

$$4({}_oF(p,n-p-1)) \leq F(p,n-p-1) \qquad \text{(III-28)}$$

where ${}_oF(p, n-p-1)$ is taken from the tabulated data for α

probability. The probability that ${}_{o}F(p, n-p-1) \leq F(p, n-p-1)$ when the dependent variable is independent of the independent variables is $1-\alpha$.

Some indication of the significance of a regression can also be obtained from confidence intervals. These indicate the range of values within which given characteristics of the data or the regression have a probability of α of being found [17].

To calculate a confidence interval for a characteristic of a data set or regression, an estimate of the variance of the characteristic is needed. If

$$\underset{\sim}{C} \equiv (\underset{\sim}{X}^T\underset{\sim}{X})^{-1} \tag{III-29a}$$

$$\underset{\sim}{C} = \begin{bmatrix} c_{oo} & \cdots & c_{op} \\ \vdots & & \vdots \\ c_{po} & \cdots & c_{pp} \end{bmatrix} \tag{III-29b}$$

and

$$s^2 \equiv MSE \tag{III-30}$$

then the variance-covariance matrix for $\underset{\sim}{B}$ is

$$\underset{\sim}{K} = \underset{\sim}{C}\, s^2 \tag{III-31a}$$

$$\underset{\sim}{K} = \begin{bmatrix} k_{oo} & \cdots & k_{op} \\ \vdots & & \vdots \\ k_{po} & \cdots & k_{pp} \end{bmatrix} \tag{III-31b}$$

where: k_{jj} = variance of b_j

k_{ij}, k_{ji} = covariance of b_i and b_j

Confidence limits for each β_j are given by

$$\beta_{jmin} = b_j - t(n-p-1)(k_{jj})^{1/2} \qquad \text{(III-32a)}$$

$$\beta_{jmax} = b_j + t(n-p-1)(k_{jj})^{1/2} \qquad \text{(III-32b)}$$

where: β_{jmin} = minimum value of β_j in interval

β_{jmax} = maximum value of β_j in interval

The value of t(n-p-1) is selected from the tabulated data for $.5+\alpha/2$ probability where α is the probability that β_j actually lies within the specified limits.

Because the β_j's are interrelated, the intervals given by (III-32) must be interpreted with care if $p > 2$. A more meaningful concept is a confidence space composed of any $\underset{\sim}{\beta}$ which satisfies

$$(\underset{\sim}{\beta}-\underset{\sim}{B})^T \underset{\sim}{X}^T \underset{\sim}{X} (\underset{\sim}{\beta}-\underset{\sim}{B}) \leq (p+1)s^2 F(p+1, n-p-1) \qquad \text{(III-33)}$$

The value of F(p+1, n-p-1) is picked from the tabulated data for α probability where α is the probability that the actual $\underset{\sim}{\beta}$ is within the specified space. Although a confidence space for $\underset{\sim}{\beta}$ is more meaningful than confidence intervals for the β_j's, an infinite number of values are required to completely specify a confidence space, hence, confidence intervals are more widely used.

Even though (III-27) may indicate that a regression is significant, one or more independent variables may be insignificant to the regression. The probability that an independent variable is significant is the probability that the corresponding β_j is nonzero. If the inequality

$$t(n-p-1) \leq |b_j| / (k_{jj})^{1/2} \qquad \text{(III-34)}$$

is true, and if t(n-p-1) is selected from the tabulated data for $.5+\alpha/2$ probability, then β_j has a probability of at least α of being nonzero. Hence, (III-34) can be used to test that the jth indpendent variable is significant at the α confidence level.

If $\underset{\sim}{X}_o$ is a row matrix containing a set of values of the independent variables, the variance for the resulting estimate of the dependent variable is

$$\sigma^2_{\hat{y}_o} = \underset{\sim}{X}_o \underset{\sim}{K} \underset{\sim}{X}_o^T \qquad \text{(III-35)}$$

where: $\sigma^2_{\hat{y}_o}$ = the variance of the estimate of the dependent variable corresponding to $\underset{\sim}{X}_o$

$\underset{\sim}{X}_o = [x_{oo} \cdots x_{po}]$

x_{io} = the value of the ith independent variable in $\underset{\sim}{X}_o$

The value of $\sigma^2_{\hat{y}_o}$ can be used to calculate a confidence interval for the estimated value of the dependent variable.

The limits on the interval are

$$\hat{\psi}_{omin} = \hat{y}_o - t(n-p-1)\sigma_{\hat{y}_o} \qquad \text{(III-36a)}$$

$$\hat{\psi}_{omax} = \hat{y}_o + t(n-p-1)\sigma_{\hat{y}_o} \qquad \text{(III-36b)}$$

where: $\hat{\psi}_{omin}$ = minimum estimated value of $\hat{\psi}_o$ in interval

$\hat{\psi}_{omax}$ = maximum estimated value of $\hat{\psi}_o$ in interval

y_o = regression estimate of the dependent variable for $\underset{\sim}{X}_o$

$\hat{\psi}_o$ = correct value of the estimated value of the dependent variable for $\underset{\sim}{X}_o$

The value of t(n-p-1) is selected from the tabulated data for $.5+\alpha/2$ probability where α is the probability $\hat{\psi}_o$ actually lies within the interval.

The variance of y_o, the actual value of the dependent variable corresponding to $\underset{\sim}{X}_o$, is

$$\sigma_{y_o}^2 = s^2 + \sigma_{\hat{y}_o}^2 \qquad \text{(III-37)}$$

where: $\sigma_{y_o}^2$ = variance of actual value of dependent variable corresponding to $\underset{\sim}{X}_o$

σ_{y_o} = standard deviation of actual value of dependent variable corresponding to $\underset{\sim}{X}_o$

This can be used to obtain the confidence limits,

$$y_{omin} = \hat{y}_o - t(n-p-1)\sigma_{y_o} \qquad \text{(III-38a)}$$

$$y_{o_{max}} = \hat{y}_o + t(n-p-1)\sigma_{y_o} \qquad \text{(III-38b)}$$

where: $y_{o_{min}}$ = minimum estimated value of y_o in interval

$y_{o_{max}}$ = maximum estimated value of y_o in interval

The value of t(n-p-1) is selected from tabulated data for $.5+\alpha/2$ probability, where α is the probability that y_o actually lies within the interval.

Optimal Parameter Selection

Often, it is desirable to pick the optimal subset of r independent variables for a regression out of a set of p variables. One definition of the optimal r variables to keep in the regression would be those variables that produce the smallest value of SSE. The most straightforward method of selecting the variables would be to perform all possible r variable regressions and select the one with the smallest SSE, but even on high speed computers this can be prohibitive in terms of computational time. Hocking and Leslie [18] and LaMotte and Hocking [19] discuss a method used to obtain the optimal regression with less computational effort. The method they discuss will be referred to here as the H.L.L. method.

The first step in the H.L.L. method is to perform the full p variable regression, retaining the values for $\underset{\sim}{B}$ and $\underset{\sim}{C}$. These values are used later.

The second step in the method is to calculate the reduction in SSE due to each variable. This is the difference between SSE produced by a regression without the variable and SSE produced by a regression including the variable. This value is given by

$$R(i) = b_i^2 / c_{ii} \qquad \text{(III-39)}$$

where: $R(i)$ = reduction in SSE produced by adding ith variable to regression

i = number of variable being tested

The third step of the method is to renumber the variables such that $R(1) \leq R(2) \leq \ldots \leq R(p)$. Because the numbering is arbitrary, this produces no loss of generality. For the remainder of this chapter, variable numbers will refer to the renumbered system.

The fourth step of the H.L.L. method uses an algorithm that requires the computation of the reduction in SSE due to groups of q variables where $q = p-r$. This can be done by an extension of (III-39). If the q variables are $i_1, \ldots, i_q$ where $i_1 < \ldots < i_q$ and if the definitions

$$\underset{\sim}{B}_1 \equiv [b_{i_1} \cdots b_{i_q}]^T \qquad \text{(III-40)}$$

where: $\underset{\sim}{B}_1$ = column matrix of elements from $\underset{\sim}{B}$ corresponding to $i_1, \ldots, i_q$

$i_1, \ldots, i_q$ = variables for which the reduction in SSE is being obtained

and

$$\underset{\sim}{C}_{11} = \begin{bmatrix} c_{i_1 i_1} & \cdots & c_{i_1 i_q} \\ \vdots & & \vdots \\ c_{i_q i_1} & \cdots & c_{i_q i_q} \end{bmatrix} \qquad \text{(III-41)}$$

where: $\underset{\sim}{C}_{11}$ = matrix of elements of $\underset{\sim}{C}$ corresponding to variables $i_1, \ldots, i_q$

are made, then the reduction in SSE due to variables $i_1, \ldots, i_q$ is

$$R(i_1, \cdots, i_q) = \underset{\sim}{B}_1^T \underset{\sim}{C}_{11}^{-1} \underset{\sim}{B}_1 \qquad \text{(III-42)}$$

where: $R(i_1, \ldots, i_q)$ = reduction in SSE produced by adding variables $i_1, \ldots, i_q$ to regression

The algorithm used in the fourth step of the H.L.L. method determines the group of q variables producing the smallest value of $R(i_1, \ldots, i_q)$. It is illustrated by the block diagram of figure III-3.

Task 1 initializes an index that indicates which

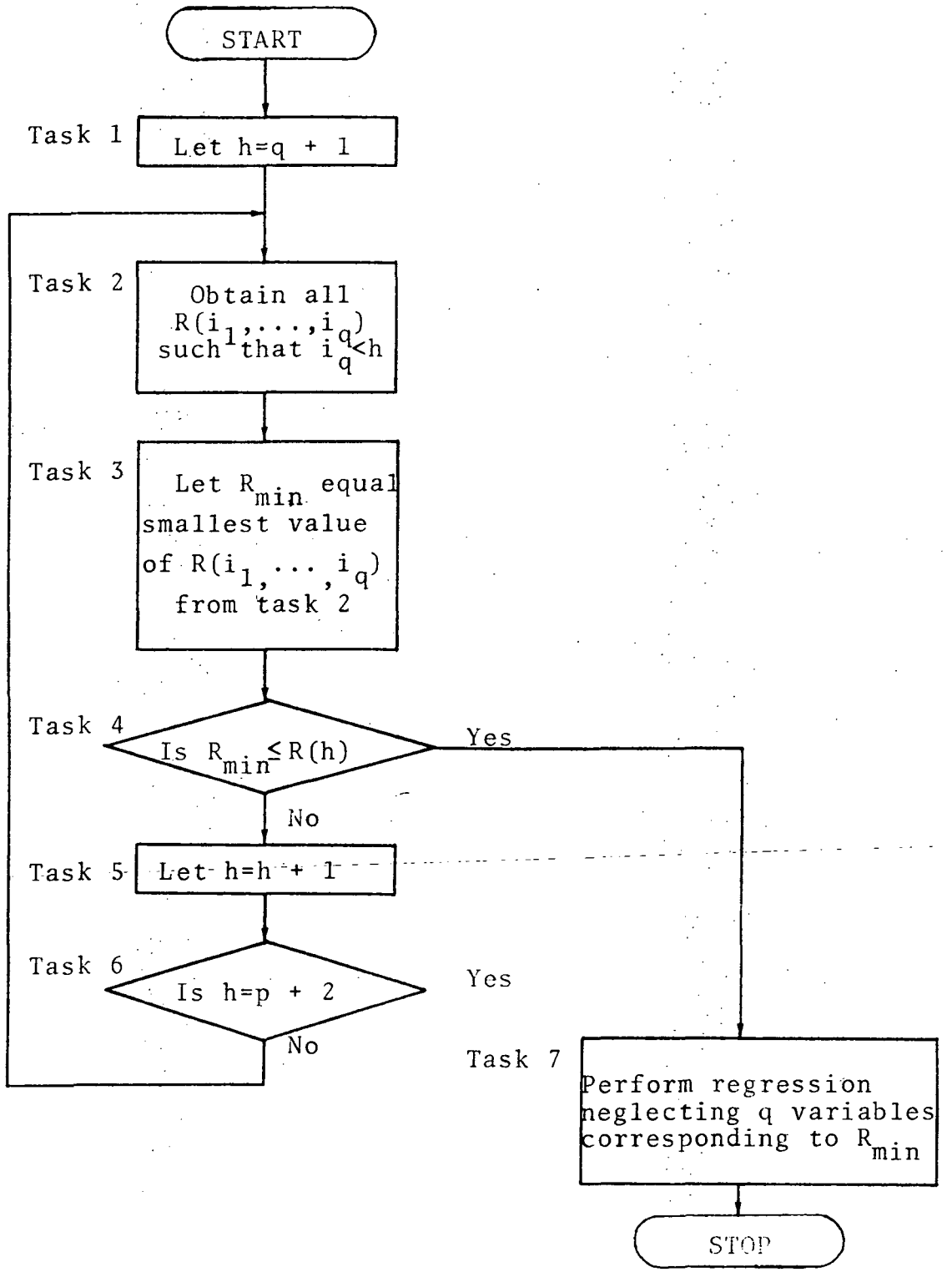


Fig. III-3 Algorithm Block Diagram

groups of q variables are to be considered during the first iteration.

In task 2, the values of $R(i_1, \ldots, i_q)$ for all groups of q variables with $i_q < h$ are obtained. Only values for groups with $i_q = h-1$ need to be computed because there are no other groups or values for other groups were computed during previous iterations.

Task 3 sets R_{min} to the smallest value of $R(i_1,\ldots,i_q)$ obtained in task 2.

Task 4 reduces the computation required by the H.L.L. method. If $R_{min} \leq R(h)$, it is the smallest possible reduction because the reduction due to any group of variables must be greater than or equal to the reduction due to any one of the variables alone, and $R(h) \leq R(H)$ if $h < H$. If $R_{min} \leq R(h)$, no further reductions need to be calculated, tasks 5 and 6 are omitted, and the process proceeds immediately to task 7, computation of the regression, neglecting the variables corresponding to R_{min}.

Task 5 increments the index to indicate the new upper bound on the variable numbers to be considered during the next iteration.

Task 6 checks if h=p+2. If it does, all possible combinations of q variables have been considered, the set producing the minimum reduction is the set corresponding

to R_{min}, and the process proceeds to task 7. If $h \neq p+2$, the process returns to task 2 for the next iteration.

Task 7 is computation of the desired regression which neglects the q variables corresponding to R_{min}.

Conclusion

This chapter describes a method of selecting the r variables from a set of p variables ($p \geq r$) that will give the optimal (smallest SSE) regression for a dependent variable. The method may be considered in six steps. First, select p independent variables that might influence or vary with the dependent variable. Second, collect values of the dependent and independent variables for at least p+1 data points. Third, select a mathematical model relating the dependent and independent variables and linearize it. Fourth, use the H.L.L. method to select and compute the optimal r variable regression. Sixth, use the various tests to determine the acceptability of the regression.

CHAPTER IV

EXPERIMENT

Rationale

Water quality data from the Houston Ship Channel was selected for testing and illustration of the techniques discussed in chapter III. Considerable concern about the quality of water in the channel has been expressed and the channel has been the subject of on-going research that includes collection of photographic and multispectral scanner data. The large number of diversified types of industry along the channel leads to a wide range and variety of water quality characteristics.

Ground Observations

The airborne data, consisting of color aerial photography and twelve channels of multispectral scanner data, were collected along the Houston Ship Channel, 12 November 1971, by a University of Michigan aircraft. The flight line originated near the mouth of the channel, approximately latitude 29°45'55", longitude 95°07'50", and terminated toward the west end, approximately latitude 29°45'15", longitude 95°15'55".

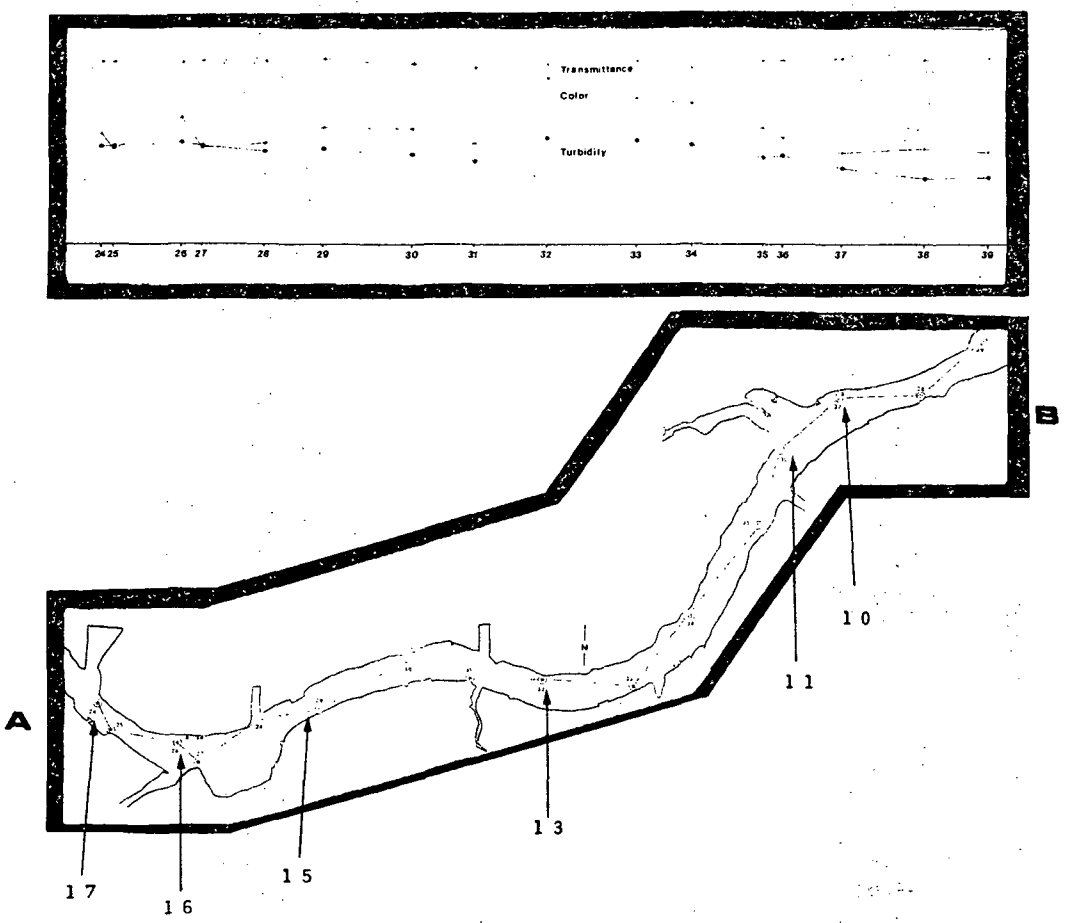
Ground observation data were collected along the flight line during a period of time three hours before to

one hour after the 12 o'clock flight. The sampling platform for the ground observation was the RV/EXCELLENCE, a water quality research vessel operated by the Civil Engineering Department of Texas A&M University.

Procedure

The RV/EXCELLENCE was equipped with various laboratory facilities for water quality measurements. One feature was an underway sampling system which pumped water from an inlet near the bow to an internal manifold system. Probes inserted through the manifold system into the water stream measured temperature, dissolved oxygen, and salinity. An outlet in the manifold system facilitated collection of water samples. The time delay between intake and sampling of the water was approximately 30 seconds.

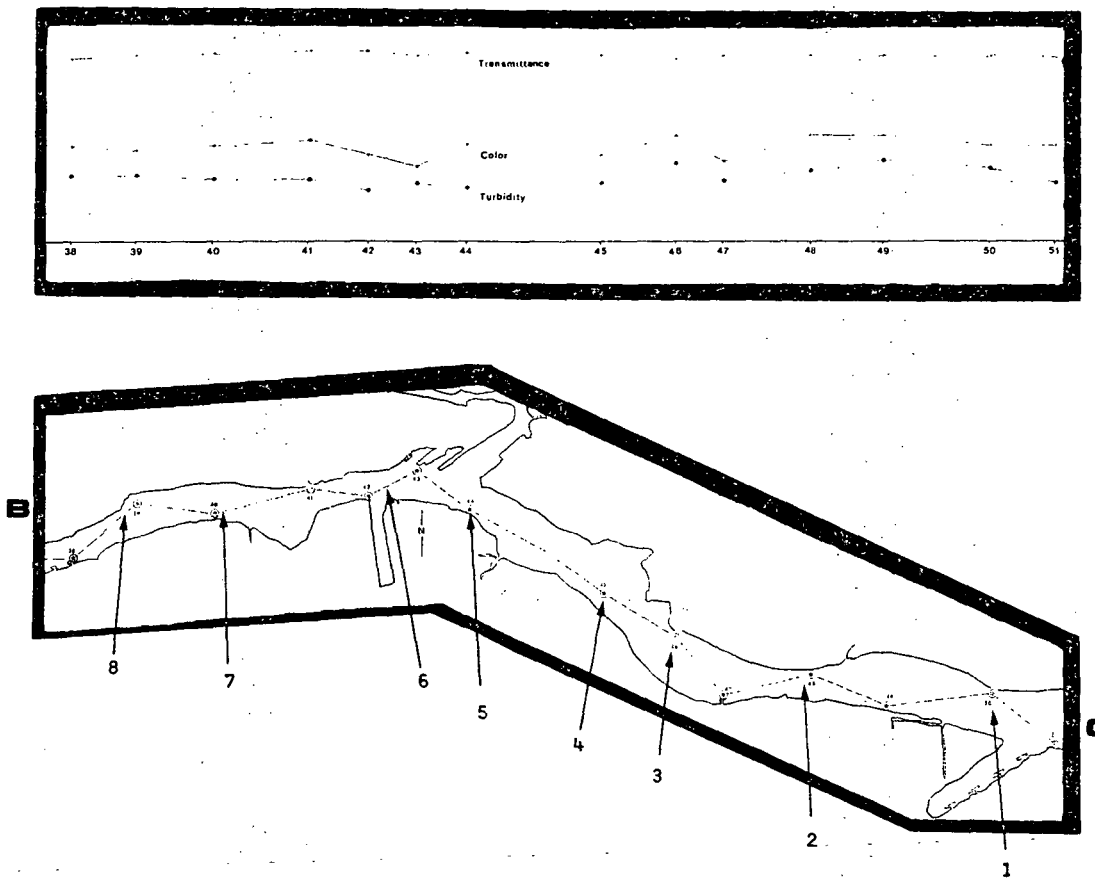
Sampling locations were selected before the flight and were located to provide good coverage of the channel. Position fixes of the vessel for the down channel trip were on alternate sides of the channel. In addition to the samples collected at these locations, or fixes, samples were also collected midway between them. These sample locations were designated by numbers equal to 0.5 plus the fix number of the immediately preceeding fix. For example, the sample location between fixes 27 and 28 was designated fix 27.5. Figures IV-1a and IV-1b show the fix locations



A is at approximately latitude 29° 43.5' longitude 95° 11.0'

B is at approximately latitude 29° 45.0' longitude 95° 11.0'

Fig. IV-1a West Half of Houston Ship Channel



B is at approximately latitude 29° 45.0' longitude 95° 11.0'

C is at approximately latitude 29° 44.0' longitude 95° 07.5'

Fig. IV-1b East Half of Houston Ship Channel

that were actually used.

The ground truth data were collected in two parts. Before the flight, the RV/EXCELLENCE moved westward along the flight line making frequent stops to collect bottom samples in conjunction with another study. During these stops, Secchi disk depth measurements were made and dye samples were released at selected locations.

After the aircraft overflight, the RV/EXCELLENCE moved eastward along the flight line measuring temperature, dissolved oxygen, and salinity. In addition, water samples were collected for tests of turbidity, transmittance, color, suspended solids, and dissolved solids. The techniques used to measure the suspended and dissolved solids are described in Standard Methods [22]. As each sample was collected, the fix and time were recorded. Tables IV-1 and IV-2 give the fix coordinates in the state plane coordinate system and the time for each fix.

The Secchi disk depth measurements consisted of lowering a standard, 30 centimeter diameter, white disk into the water and recording the depth at which it disappeared. It has been established empirically that dividing the disk depth into 1.7 gives the extinction coefficient [23].

The dye experiment consisted of mixing 2.2 liters of rhodamine WT 20% dye with channel water and releasing it into the channel. This experiment was to test the

TABLE IV-1

UPSTREAM DATA

Fix No.	Time Hr:Min	State Plane Coordinates X	Y	Secchi Disk Depth (Meters)	Remarks
1	08:48	3,226,757	711,516	0.457	
2	08:59	3,222,631	711,832	0.457	Note
3	09:10	3,220,021	712,189	0.457	
4	09:14	3,218,568	713,600	0.518	Note
5	09:20	3,215,789	715,116	0.549	
6	09:26	3,213,473	715,495	0.549	Note
7	09:34	3,210,442	714,989	0.549	
8	09:38	3,208,631	714,737	0.549	Note
10	09:43	3,205,936	713,284	0.518	
11	09:47	3,204,694	712,232	0.518	Note
13	09:51	3,199,852	706,968	0.366	Note
15	10:05	3,194,652	705,895	0.305	Note
16	10:11	3,192,147	705,011	0.396	Note
17	10:18	3,190,378	705,558	0.366	

Note: 2.2 liters of dye released

TABLE IV-2

DOWNSTREAM DATA COORDINATES

FIX	TIME			UNSHIFTED STATE PLANE COORDINATES		SHIFTED STATE PLANE COORDINATES	
	HR.	MIN.	SEC.	X	Y	X	Y
24.0	11	24	0	3190526	705684	3190526	705684
24.5	11	24	30	-------	------	3190526	705684
25.0	11	25	0	3190905	705010	3190715	705347
25.5	11	28	0	-------	------	3191568	704968
26.0	11	30	0	3192231	704926	3192098	704934
26.5	11	31	0	-------	------	3192393	704847
27.0	11	32	0	3192842	704610	3192689	704689
27.5	11	33	0	-------	------	3193020	704796
28.0	11	35	0	3193915	705726	3193736	705540
28.5	11	35	30	-------	------	3193915	705726
29.0	11	37	0	3195031	705957	3194752	705899
29.5	11	37	30	-------	------	3195031	705957
30.0	11	41	30	3196821	707368	3196622	707211
30.5	11	42	0	-------	------	3196821	707368
31.0	11	45	0	3198252	707010	3198047	707061
31.5	11	45	15	-------	------	3198149	707036
32.0	11	46	50	3199663	707136	3199278	707102
32.5	11	47	15	-------	------	3199598	707131
33.0	11	50	0	3201494	706947	3201204	706977
33.5	11	50	45	-------	------	3201571	707053
34.0	11	53	45	3202652	708547	3202497	708333
34.5	11	54	0	-------	------	3202574	708440
35.0	11	57	30	3204105	710568	3203911	710293
35.5	11	57	45	-------	------	3204008	710433
36.0	12	0	30	3204505	712105	3204438	711849
36.5	12	1	30	-------	------	3204682	712300
37.0	12	4	0	3205747	713473	3205569	713278
37.5	12	4	30	-------	------	3205747	713473
38.0	12	7	30	3207578	713684	3207316	713654
38.5	12	8	0	-------	------	3207578	713684

CONTINUATION TABLE IV-2

DOWNSTREAM DATA COORDINATES

FIX	TIME			UNSHIFTED STATE PLANE COORDINATES		SHIFTED STATE PLANE COORDINATES	
	HR.	MIN.	SEC.	X	Y	X	Y
39.0	12	11	30	3209305	715073	3209089	714899
39.5	12	12	0	-------	------	3209305	715073
40.0	12	13	30	3210421	714736	3210142	714821
40.5	12	13	45	-------	------	3210281	714779
41.0	12	16	15	3212399	715368	3212039	715253
41.5	12	16	30	-------	------	3212219	715310
42.0	12	19	15	3213557	715305	3213364	715315
42.5	12	19	45	-------	------	3213557	715305
43.0	12	21	15	3214568	715894	3214315	715747
43.5	12	22	0	-------	------	3214675	715823
44.0	12	24	0	3215747	715115	3215532	715257
44.5	12	24	15	-------	------	3215650	716160
45.0	12	30	30	3218610	713536	3218365	713690
45.5	12	30	45	-------	------	3218499	713597
46.0	12	33	30	3220126	712694	3219873	712835
46.5	12	34	15	-------	------	3220194	712618
47.0	12	37	30	3221221	711473	3221084	711636
47.5	12	38	0	-------	------	3221221	711473
48.0	12	40	15	3222778	711978	3222494	711887
48.5	12	41	0	-------	------	3222950	711976
49.0	12	42	45	3224505	711452	3224159	711597
49.5	12	43	0	-------	------	3224332	711505
50.0	12	47	0	3226778	711705	3226510	711675
50.5	12	49	45	-------	------	3227456	711244
51.0	12	51	45	3228210	710736	3228059	710839

feasibility of using photography or multispectral scanner data to determine current velocity and diffusion coefficients.

Temperature, dissolved oxygen, and salinity were measured using standard probes inserted into the water distribution system. The values were recorded on chart recorders for essentially continuous records of these parameters. The chart data was related to the correct locations by marking the data for each fix with the fix number.

Turbidity, transmittance, and color were measured using an Hach DC-DR colorimeter. The basis for this device is the characteristic absorption spectrum. This means that the absorption spectrum of an impure water sample is a function of the types and quantities of the impurities. By looking at the absorption of particular bands by the sample, the quantities of certain of these impurities can be estimated.

In the HACH DC-DR colorimeter, represented in figure IV-2, the light energy from an electric bulb is filtered by an interchangeable filter, passes 2.2 cm through a water sample, and is measured by a photodetector connected to a meter with interchangeable scales. The filter is selected to pass the band most sensitive to the parameter

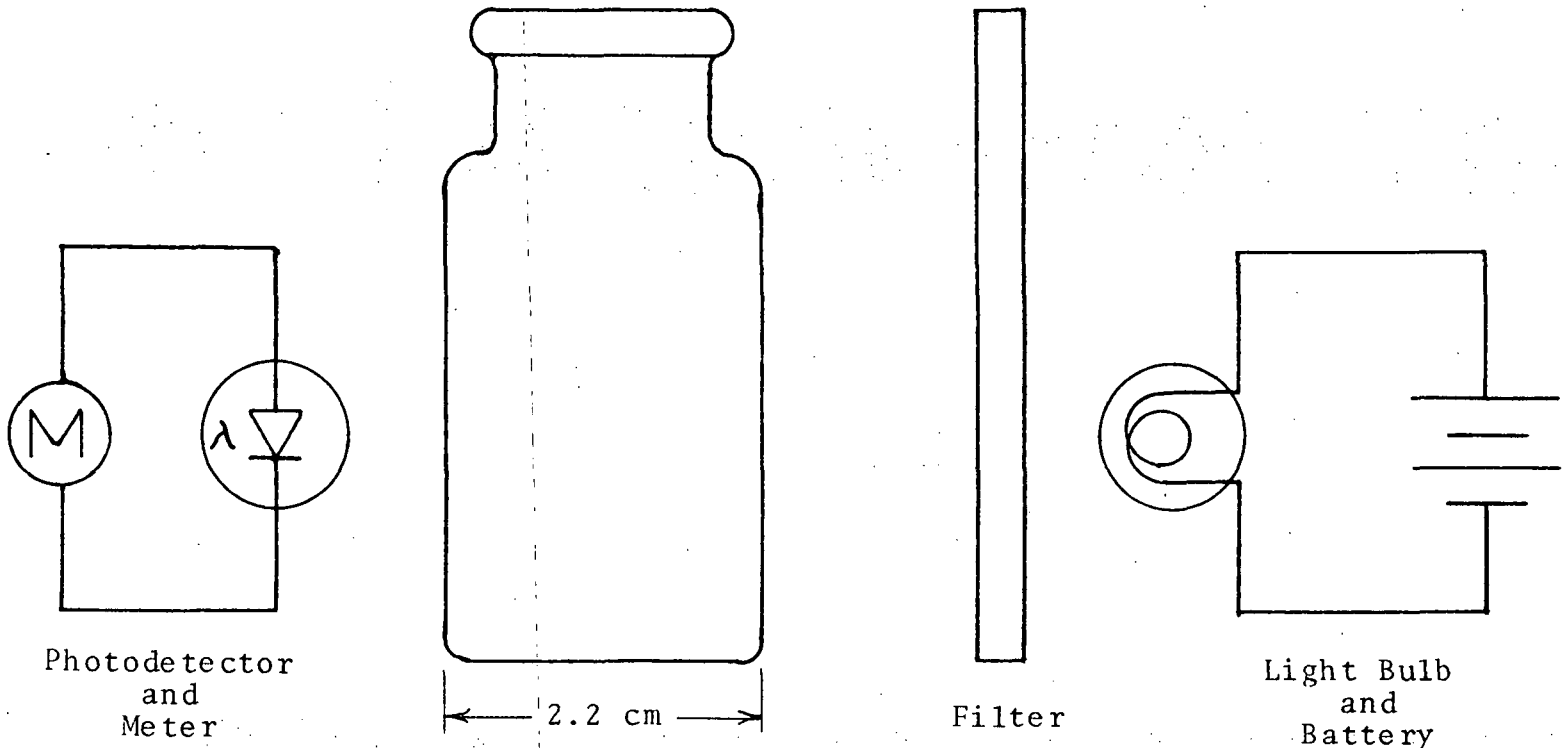


Fig. IV-2 Hach DC-DR Colorimeter

being measured, while the scale is selected to read directly in the units best suited to the measurement. In use, the correct filter and scale were selected according to a chart. They were installed, a sample of distilled water was used to zero the meter, then samples were inserted and measurements were taken.

For the tests in this study, both a large sample bottle and a colorimeter sample bottle were filled from the manifold system outlet at each sample location. The colorimeter sample was then used to measure turbidity. After all sample locations were completed, the filter and scale on the colorimeter were changed and samples were drawn from the large bottles to measure transmittance and color. To maintain accuracy, a sample of pure water was used to rezero the meter after every tenth sample. Usually, little or no adjustment was needed.

After the flight, standard laboratory tests were conducted on some of the samples to measure total suspended and dissolved solids. In addition to these tests, a few samples were tested for salinity as a check on the data recorded on the boat. The laboratory data are listed in table IV-3.

Data Processing

The raw data that were collected required some initial

TABLE IV-3

LABORATORY DATA

Fix	Total Suspended Solids (MG/L)	Dissolved Solids (MG/L)	Salinity (0/00)
24.0		17,300	
25.0	25	17,400	
26.0			14.06
26.5	25	16,000	
27.5	28	15,400	
29.0	24	16,000	
29.5			13.83
30.0	21	15,900	
31.5	23		
32.5	29		
34.0	26		
34.5			15.76
35.0	23		
36.5	23		
37.5	22		
39.0	18	17,800	
40.0	17	17,900	
41.5	21	18,200	
42.5	24	18,500	
44.0	21	18,300	
45.0	27	18,200	
46.5	22		
47.0			16.79
47.5	20		
48.0	24		
49.0	36		
50.0	28		
51.0	24		

processing to give the actual values of the various parameters. The Secchi disk depths, measured in feet, had to be converted to meters, then divided into an empirical factor of 1.7 to give the extinction coefficients. Fixes, times, state plane coordinates, Secchi disk depths, and locations of dye releases are given in table IV-1 (p. 75). Table IV-4 gives the extinction coefficients obtained by this method.

The actual values of temperature in degrees Fahrenheit were obtained by multiplying the chart values by ten. The factor for obtaining salinity values in parts per thousand from the chart values was 1.92, while dissolved oxygen values in milligrams per liter were read directly from the chart. Turbidity in Jackson turbidity units, transmittance in percent, and color on the APHA platinum-cobalt standard were read directly from the colorimeter. Table IV-5 lists the actual values of all of these parameters.

Table IV-3 (p. 81) gives the parameter values that were obtained in the laboratory. The suspended and dissolved solids were measured in milligrams per liter, while the salinity was measured in parts per thousand.

Coordinate Compensation

In order to compare the ground truth data to the

TABLE IV-4

EXTINCTION COEFFICIENTS

Fix (Upstream)	Extinction Coefficient (Secchi Disk)	Fix (Down Stream)	Extinction Coefficient (Transmittance)
1.0	3.72	50.0	3.54
2.0	3.72	48.0	3.79
3.0	3.72	46.0	4.04
4.0	3.28	45.0	3.54
5.0	3.10	44.0	3.05
6.0	3.10	42.0	2.09
7.0	3.10	40.0	3.05
8.0	3.10	38.5	3.79
10.0	3.28	37.0	3.30
11.0	3.28	36.0	4.04
13.0	4.65	32.0	5.04
15.0	5.56	28.5	5.55
16.0	4.28	26.0	4.79
17.0	4.65	24.5	4.04

TABLE IV-5

CHART AND COLORIMETER DATA

FIX	TEMPERATURE (DEG F)	OXYGEN (MG/L)	SALINITY (0/00)	TURBIDITY (JTU)	TRANSMITTANCE (PERCENT)	COLOR
24.0	75.5C	----	12.86	48.50	91.50	110.00
24.5	75.5C	----	12.86	48.50	91.50	95.00
25.0	75.5C	----	12.96	49.00	91.50	95.00
25.5	75.00	0.0	13.06	42.00	91.00	100.00
26.0	76.00	0.30	12.00	50.50	90.00	125.00
26.5	77.5C	0.0	10.66	48.50	90.50	115.00
27.0	77.00	0.0	11.52	49.00	91.00	115.00
27.5	76.00	0.0	11.62	48.00	89.50	95.00
28.0	78.00	0.0	9.22	46.00	91.00	100.00
28.5	77.5C	0.0	10.08	47.00	88.50	135.00
29.0	77.00	0.0	10.85	47.00	92.00	95.00
29.5	76.00	0.0	11.52	50.50	90.00	115.00
30.0	76.5C	0.0	11.90	45.50	90.50	115.00
30.5	76.5C	0.0	12.00	45.00	90.50	115.00
31.0	76.00	0.0	12.48	41.00	87.50	100.00
31.5	82.5C	0.0	13.92	37.00	90.50	95.00
32.0	80.00	0.15	12.29	52.00	89.50	165.00
32.5	79.5C	0.0	12.19	60.00	89.50	175.00
33.0	78.00	0.15	12.48	51.50	91.00	145.00
33.5	78.00	0.0	12.58	48.50	90.50	130.00

CONTINUATION TABLE IV-5

CHART AND COLORIMETER DATA

FIX	TEMPERATURE (DEG F)	OXYGEN (MG/L)	SALINITY (0/00)	TURBIDITY (JTU)	TRANSMITTANCE (PERCENT)	COLOR
34.0	78.0C	0.05	13.06	49.00	87.50	140.00
34.5	78.00	0.10	12.67	44.00	90.50	125.00
35.0	77.00	0.05	12.96	42.50	91.00	115.00
35.5	77.0C	0.0	12.96	39.00	88.50	95.00
36.0	77.0C	0.0	12.96	44.00	91.50	105.00
36.5	77.00	0.0	13.06	41.50	91.50	105.00
37.0	76.5C	0.65	12.00	38.00	93.00	90.00
37.5	75.5C	1.35	11.33	29.50	93.50	95.00
38.0	78.00	0.55	13.15	32.00	91.50	95.00
38.5	77.5C	0.40	13.06	31.00	92.00	85.00
39.0	77.50	1.40	12.96	32.00	93.00	90.00
39.5	78.00	2.15	12.96	33.00	91.50	85.00
40.0	77.0C	0.25	13.15	31.00	93.50	95.00
40.5	77.0C	0.25	13.25	27.00	93.00	100.00
41.0	76.50	0.35	13.25	29.50	94.50	100.00
41.5	77.0C	0.35	13.25	21.50	93.00	85.00
42.0	77.50	0.25	13.25	25.50	95.50	85.00
42.5	----	----	----	28.50	93.50	80.00
43.0	----	----	----	28.50	92.50	75.00
43.5	----	----	----	29.00	92.50	100.00

CONTINUATION TABLE IV-5

CHART AND COLORIMETER DATA

FIX	TEMPERATURE (DEG F)	OXYGEN (MG/L)	SALINITY (O/OO)	TURBIDITY (JTU)	TRANSMITTANCE (PERCENT)	COLOP
44.0	----	----	----	26.50	93.50	95.00
44.5	----	----	----	25.50	93.50	85.00
45.0	----	----	----	29.00	92.50	85.00
45.5	----	----	----	31.00	94.00	80.00
46.0	75.5C	0.15	13.34	38.50	91.50	105.00
46.5	75.5C	0.15	13.34	35.50	93.00	90.00
47.0	75.0C	0.20	13.44	30.50	93.00	80.00
47.5	75.0C	0.15	13.44	34.50	91.50	95.00
48.0	75.0C	0.15	13.44	35.00	92.00	105.00
48.5	75.0C	0.15	13.44	31.00	92.50	90.00
49.0	75.0C	0.15	13.44	40.50	93.00	105.00
49.5	75.0C	0.20	13.54	29.50	93.00	90.00
50.0	75.00	0.25	13.44	36.50	92.50	95.00
50.5	75.00	0.20	13.44	30.50	92.50	95.00
51.0	74.5C	0.30	13.44	28.50	91.50	95.00

photographic and multispectral scanner data, each water sample location was needed. The coordinates of each fix could be read from the map, but the 30 second time delay introduced by the bow sampling system meant that each sample was actually taken some distance back along the course from the corresponding fix. The sampling coordinates were obtained using the assumptions that the boat maintained a constant speed and that it followed a straight line course between fixes.

In general, coordinate compensation and midfix coordinates were obtained by applying (IV-1) to both the x and y coordinates.

$$C = R_j - \frac{(R_{j-1} - R_j)(T_j - T + 30)}{(T_{j-1} - T_j)} \qquad \text{(IV-1)}$$

where:

- C = compensated coordinate of sample location
- R_j = coordinate of fix to which the boat was going, fix j
- R_{j-1} = coordinate of fix from which the boat was coming, fix j-1
- T_j = time (sec.) boat was at fix j
- T_{j-1} = time (sec.) boat was at fix j-1
- T = time sample was taken

In some cases, $(T_j - T + 30)$ was greater than $(T_j - T_{j-1})$, that is, the delay was sufficient to shift the sample location

back past the fix from which the boat was coming. In that case, (IV-2) was used because it took account of this fact.

$$C = R_{j-1} - \frac{(R_{j-2} - R_{j-1})(T_{j-1} - T + 30)}{(T_{j-2} - T_{j-1})} \qquad \text{(IV-2)}$$

where: R_{j-2} = coordinate of fix j-2

T_{j-2} = time (sec.) boat was at fix j-2

In two cases, neither (IV-1) nor (IV-2) was applicable. The first case was fix 24. Since there was no way to know the speed or course of the boat approaching this fix, its compensated coordinates could not be found, so its coordinates were treated as though they needed no compensation.

The second case occurred between fixes 44 and 45, where channel traffic forced a change in the planned course as shown in figure IV-1b (p. 73). To obtain the compensated coordinates, the map distance between fixes 44 and 45, along the course shown, was found. This distance was then divided by the time it took to go from fix 44 to fix 45 to give the scale speed, S. Equation (IV-3) gives the scale distance from fix 44 to fix 45.

$$D = S(T - T_{44} - 30) \qquad \text{(IV-3)}$$

where: D = scale distance along the course from fix 44 to actual location from which sample for fix 44.5 (45) was taken

T_{44} = time (sec.) sample for fix 44 was taken

T = time (sec.) sample for fix 44.5 (45) was taken

S = scale speed of boat between fixes 44 and 45

Fix 44.5 shifted back past fix 44, hence (IV-2) was used.

Data Analysis

Several steps were taken to enhance the data and to check its accuracy. To enhance the data, the means and variances for the temperature, dissolved oxygen, salinity, turbidity, transmittance, and color were calculated (table IV-6). It can be observed that several of the parameters, such as temperature and transmittance, are large compared to their variance. In order to make the variation more obvious, the variations from the mean were also calculated (table IV-7).

Two sets of data, the transmittance and the salinity, were readily susceptible to checks. Transmittance, related to the extinction coefficient by (IV-4), could be checked by comparing the values of extinction coefficient obtained by (IV-4) with those obtained from the Secchi disk depths.

TABLE IV-6

CHART AND COLORIMETER DATA STATISTICS

	MEAN	VARIANCE
TEMPERATURE	7.669E 01	2.336E 00
OXYGEN	2.400E-01	1.747E-01
SALINITY	1.263E 01	9.531E-01
TURBIDITY	3.843E 01	8.261E 01
TRANSMITTANCE	9.162E 01	2.852E 00
COLOR	9.745E 01	9.656E 02

TABLE IV-7

CHART AND COLORIMETER DATA VARIATION

FIX	TEMPERATURE (DEG F)	OXYGEN (MG/L)	SALINITY (0/00)	TURBIDITY (JTU)	TRANSMITTANCE (PERCENT)	COLOR
24.0	-1.19	----	0.23	10.07	-0.12	12.55
24.5	-1.19	----	0.23	10.07	-0.12	-2.45
25.0	-1.19	----	0.33	10.57	-0.12	-2.45
25.5	-1.69	-0.24	0.43	3.57	-0.62	2.55
26.0	-0.69	0.06	-0.63	12.07	-1.62	27.55
26.5	0.81	-0.24	-1.97	10.07	-1.12	17.55
27.0	0.31	-0.24	-1.11	10.57	-0.62	17.55
27.5	-0.69	-0.24	-1.01	9.57	-2.12	-2.45
28.0	1.31	-0.24	-3.41	7.57	-0.62	2.55
28.5	0.81	-0.24	-2.55	8.57	-3.12	37.55
29.0	0.31	-0.24	-1.78	8.57	0.38	-2.45
29.5	-0.69	-0.24	-1.11	12.07	-1.62	17.55
30.0	-0.19	-0.24	-0.73	7.07	-1.12	17.55
30.5	-0.19	-0.24	-0.63	6.57	-1.12	17.55
31.0	-0.69	-0.24	-0.15	2.57	-4.12	2.55
31.5	5.81	-0.24	1.29	-1.43	-1.12	-2.45
32.0	3.31	-0.09	-0.34	13.57	-2.12	67.55
32.5	2.81	-0.24	-0.44	21.57	-3.12	77.55
33.0	1.31	-0.09	-0.15	13.07	-0.62	47.55
33.5	1.31	-0.24	-0.05	10.07	-1.12	32.55

TABLE IV-7

CHART AND COLORIMETER DATA VARIATION

FIX	TEMPERATURE (DEG F)	OXYGEN (MG/L)	SALINITY (0/00)	TURBIDITY (JTU)	TRANSMITTANCE (PERCENT)	COLOR
34.0	1.31	-0.19	0.43	10.57	-4.12	42.55
34.5	1.31	-0.14	0.04	5.57	-1.12	27.55
35.0	0.31	-0.19	0.33	4.07	-0.62	17.55
35.5	0.31	-0.24	0.33	0.57	-3.12	-2.45
36.0	0.31	-0.24	0.33	5.57	-0.12	7.55
36.5	0.31	-0.24	0.43	3.07	-0.12	7.55
37.0	-0.19	0.41	-0.63	-0.43	1.38	-7.45
37.5	-1.19	1.11	-1.30	-8.93	1.88	-2.45
38.0	1.31	0.31	0.52	-6.43	-0.12	-2.45
38.5	0.81	0.16	0.43	-7.43	0.38	-12.45
39.0	0.81	1.16	0.33	-6.43	1.38	-7.45
39.5	1.31	1.91	0.33	-5.43	-0.12	-12.45
40.0	0.31	0.01	0.52	-7.43	1.88	-2.45
40.5	0.31	0.01	0.62	-11.43	1.38	2.55
41.0	-0.19	0.11	0.62	-8.93	2.88	2.55
41.5	0.31	0.11	0.62	-16.93	1.38	-12.45
42.0	0.81	0.01	0.62	-12.93	3.88	-12.45
42.5	----	----	----	-9.93	1.88	-17.45
43.0	----	----	----	-9.93	0.88	-22.45
43.5	----	----	----	-9.43	0.88	2.55

TABLE IV-7

CHART AND COLORIMETER DATA VARIATION

FIX	TEMPERATURE (DEG F)	OXYGEN (MG/L)	SALINITY (O/OO)	TURBIDITY (JTU)	TRANSMITTANCE (PERCENT)	COLOR
44.0	----	----	----	-11.93	1.88	-2.45
44.5	----	----	----	-12.93	1.88	-12.45
45.0	----	----	----	-9.43	0.88	-12.45
45.5	----	----	----	-7.43	2.38	-17.45
46.0	-1.19	-0.09	0.71	0.07	-0.12	7.55
46.5	-1.19	-0.09	0.71	-2.93	1.38	-7.45
47.0	-1.69	-0.04	0.81	-7.93	1.38	-17.45
47.5	-1.69	-0.09	0.81	-3.93	-0.12	-12.45
48.0	-1.69	-0.09	0.81	-3.43	0.38	7.55
48.5	-1.69	-0.09	0.81	-7.43	0.88	-7.45
49.0	-1.69	-0.09	0.81	2.07	1.38	7.55
49.5	-1.69	-0.04	0.91	-8.93	1.38	-7.45
50.0	-1.69	0.01	0.81	-1.93	0.88	-2.45
50.5	-1.69	-0.04	0.81	-7.93	0.88	-2.45
51.0	-2.19	0.06	0.81	-9.93	-0.12	-2.45

$$k = -\ln\left(\frac{J}{D}\right) \qquad \text{(IV-4)}$$

where: k = extinction coefficient (m^{-1})

J = transmittance x 100

D = distance (m) the light travels in the water

The values obtained by both methods are given in table IV-4 (p. 83), where the downstream fixes located closest to the upstream fixes were used. The table shows that the two sets of data are virtually the same.

The salinity values obtained from the chart data, table IV-5 (p. 84) were checked by comparing the salinity values obtained in the laboratory (table IV-3 (p. 81)), to them. The laboratory values were consistently 2-3 parts per thousand higher. One explanation of this is that during the several days between the time the samples were collected and the time the laboratory measurements were made, evaporation of water from the sample bottles increased the salinity.

Photography

Due to the effort and cost of processing, only three of the photo-transparencies were processed. Of these three

transparencies, the first viewed fix 44, the second viewed fix 31.5, and the third viewed fixes 26, 26.5 and 27.

Quantitative data was obtained from the photography by having portions of the three 9 x 9 inch photo-transparencies densitized by Optronics International Incorporated. The data was recorded on digital magnetic tape as numbers from 0 to 255. These values represented film densities from 0 to 2.55. The data was densitized in three bands; blue, green and red. Each data cell was represented by three data values, one for each band.

The resolution initially available from the photography was approximately 1.25 by 1.25 feet. This was greater than was necessary, therefore, to reduce processing time and data storage requirements, it was reduced to approximately 10 x 10 feet. For the reduction, areas originally represented by sixty four cells, eight cells on a side, were represented by a single data cell. The band values for the cell were obtained by averaging the original sixteen center cells.

After the data was reduced, an eight-level gray scale printout of the data for one band of each transparency was produced by computer. This made it possible to select the data corresponding to each fix location.

Scanner Data

The multispectral scanner used in this study had a 90 degree unobscured scan field of view. The scanner had a nominal spatial resolution of 3 milliradians, a nominal temperature resolution of .25°K, and a nominal radiation reflection resolution of 1%. References for calibration consisted of sources within the range .3 to 14 microns, plus sky illumination reference.

The scanner had three detector positions. The first position had a HgCdTe detector nominally filtered for 9.3 to 11.7 micron wavelengths. The second position had a three element InAs detector with the elements nominally filtered for 1.0 to 1.4, 1.5 to 1.8, and 2.0 to 2.6 micron wavelengths. The third position had a prism spectrometer with nine bands, nominally .41 to .48, .46 to .49, .48 to .52, .50 to .54, .53 to .57, .55 to .60, .58 to .64, .62 to .69, and .67 to .91 micron wavelengths. Due to lack of channels on the data recorder, only eight bands were used. The ninth, .48 to .52 microns, was neglected [14].

Analog Data Description

The scanner data were recorded in analog form at 60 inches per second on 14 channel magnetic tape. All of the channels except channel 7 were recorded using frequency

modulation with a center frequency of 216 KHz and a deviation of ±30 percent. A positive deviation indicated a darker (or colder) area.

Twelve of the 14 recorded channels contained spectral data, the other two contained synchronization data. Table IV-8 lists the bands for each channel. The information bearing portion of the scanner data for each scan line was collected during the time the scanner mirror viewed the ground through the aperture. The remainder of the data for each scan contained calibration pulses. Figure IV-3 shows typical synchronization and scan data for one scan line.

Channel 7, the only channel that was recorded directly instead of using frequency modulation, and channel 14 were the synchronization channels. Their purpose was to indicate the start of each scan line. For each scan line, channel 14 contained a synchronization pulse, a start of scan pulse, and a series of pulses indicating the number of the scan line.

Initial Scanner Data Processing

The analog data was contained on two magnetic tapes, each of which contained data for several flight lines. Only two flight lines on each tape, lines 7 and 8, were

TABLE IV-8

SPECTRAL BANDS REPRESENTED BY EACH DATA CHANNEL

Channel	Band	Band Description
1	0.41-0.48	Visible
2	0.46-0.49	Visible
3	0.50-0.54	Visible
4	0.52-0.57	Visible
5	0.55-0.60	Visible
6	0.58-0.64	Visible
7	---	Synchronization Data
8	0.67-0.94	Visible to Near Infrared
9	1.00-1.40	Near Infrared
10	1.50-1.80	Near Infrared
11	2.00-2.60	Near Infrared to Thermal Infrared
12	9.30-11.50	Thermal Infrared
13	0.63-0.70	Visible
14	---	Synchronization Data

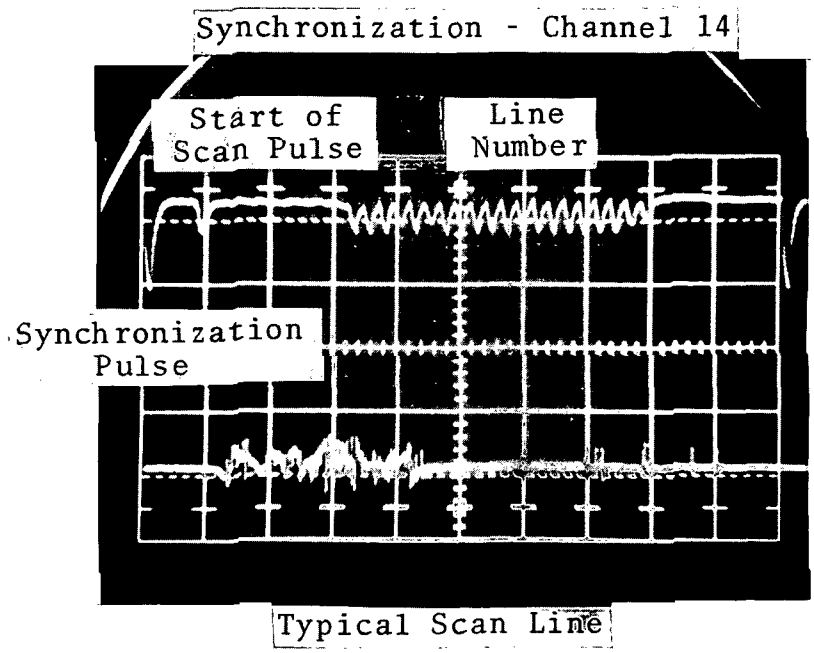


Fig. IV-3 Typical Scanner Data

of interest. To locate the data for the desired flight lines, the network shown in figure IV-4 was used to display the data on an oscilloscope in a form much like a television picture. Photographs of the oscilloscope CRT were than taken and used to form mosaics as shown in figures IV-5, IV-6a and IV-6b. Comparison of the mosaics with maps and photographs of the ship channel indicated the location of the desired data.

During display of the flight lines on tape two, difficulty was encountered with synchronization and it was impossible to obtain a reliable trigger signal. Since these were repetitions of the flight lines on tape one, tape two was disregarded.

The only 1 inch drive recorder available for the analog tapes was incompatible with the equipment needed to digitize the data, therefore, the desired data was reproduced using a 1/4 inch drive four channel recorder and the network of figure IV-7. Only two of the four channels on the second recorder could be used to record data. One channel was inoperative, and the other channel was required to record a "trace pulse" for synchronization. To reproduce more than two data channels required more "passes" through the original tape. Because of the difficulty involved in reproducing the data, only those data channels expected to add significantly more infor-

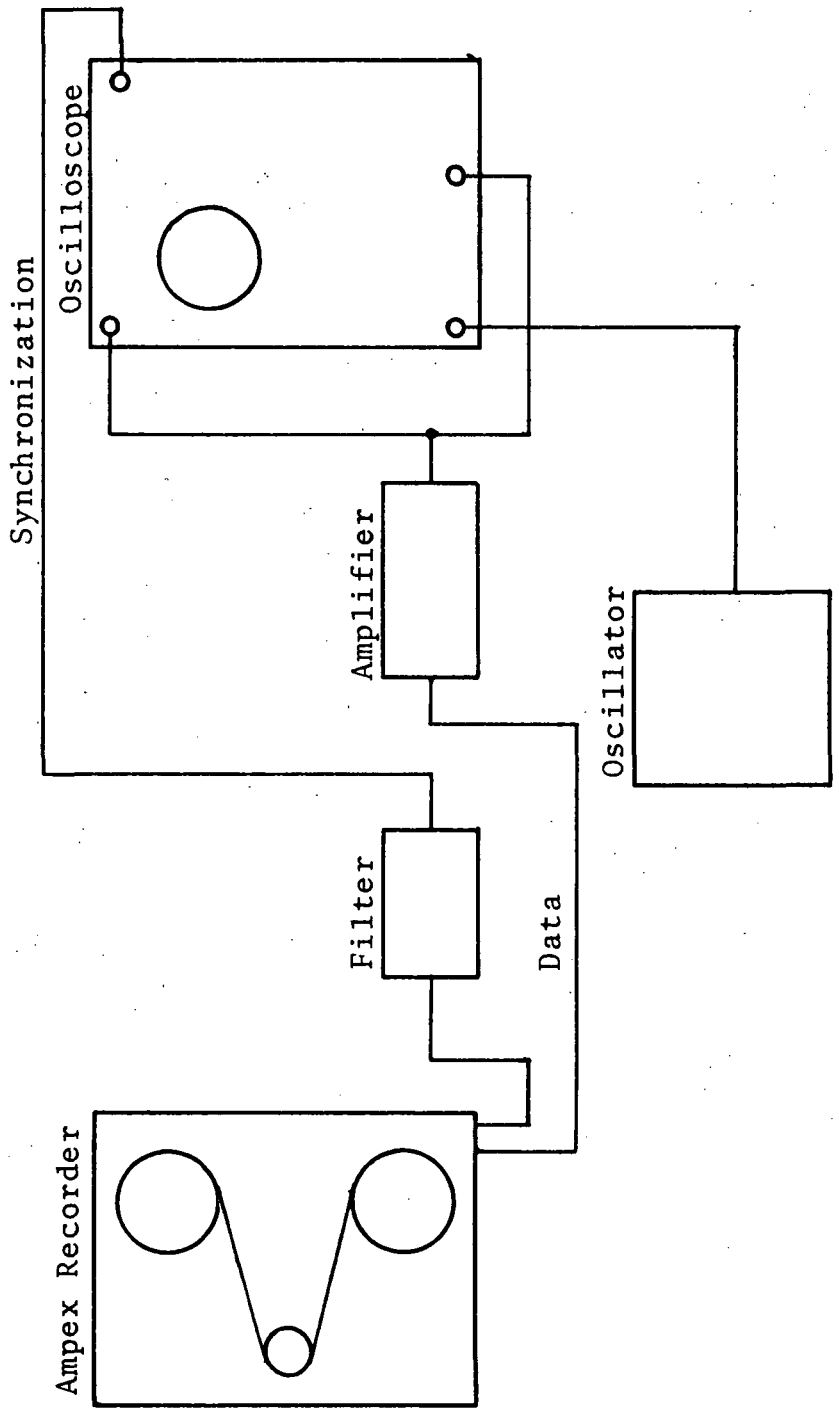


Fig. IV-4 Monitoring System

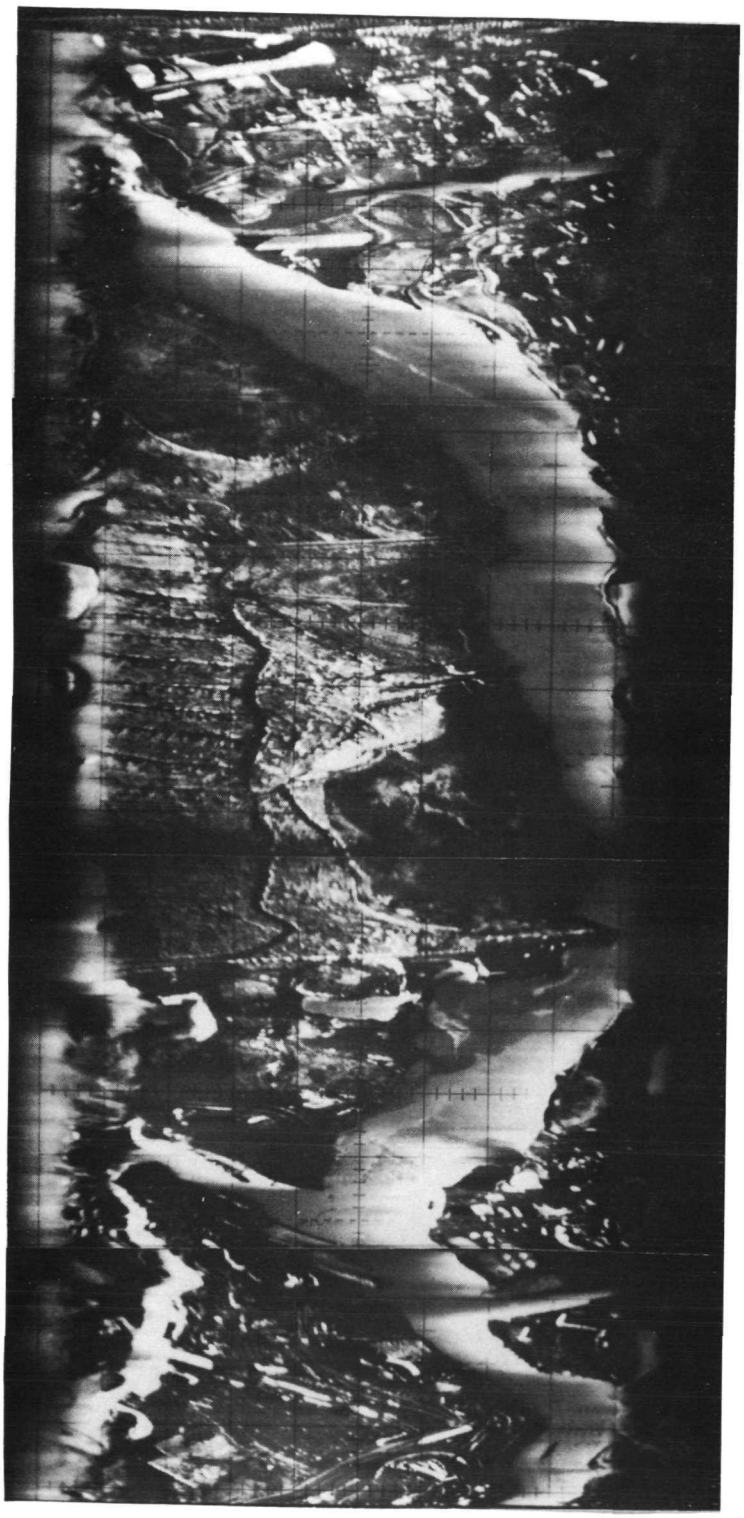


Fig. IV-5 Mosaic of Flight Line 7

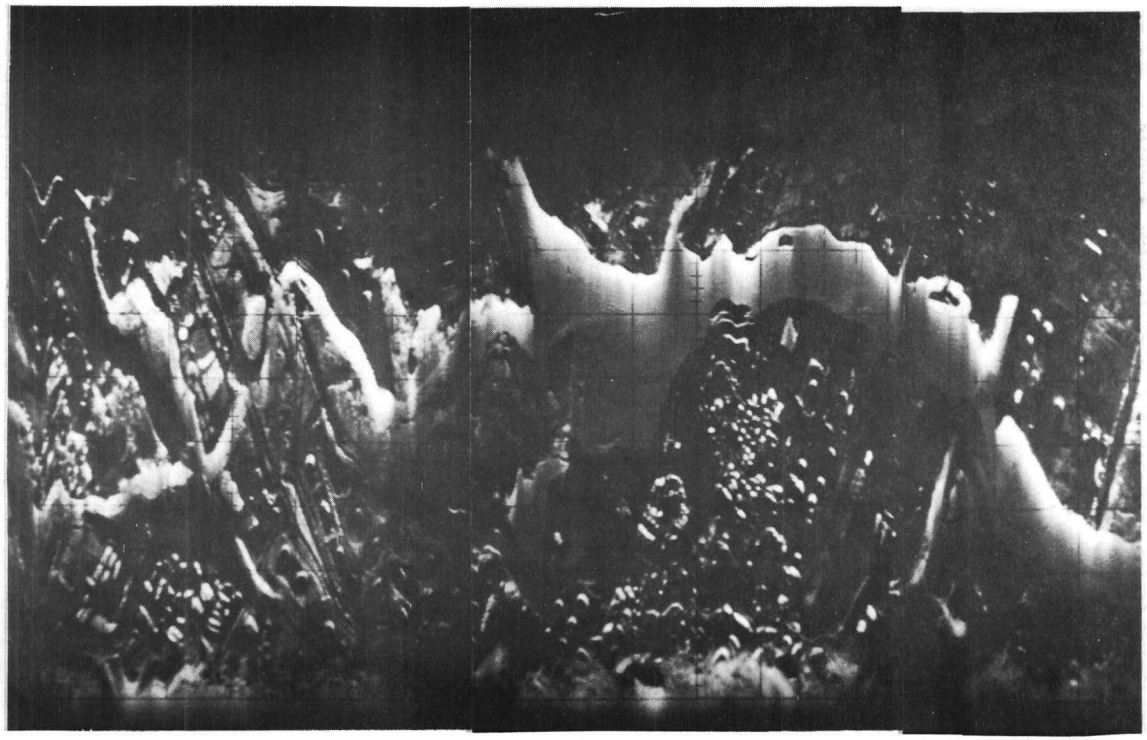


Fig. IV-6a Mosaic of West Half of Flight Line 8

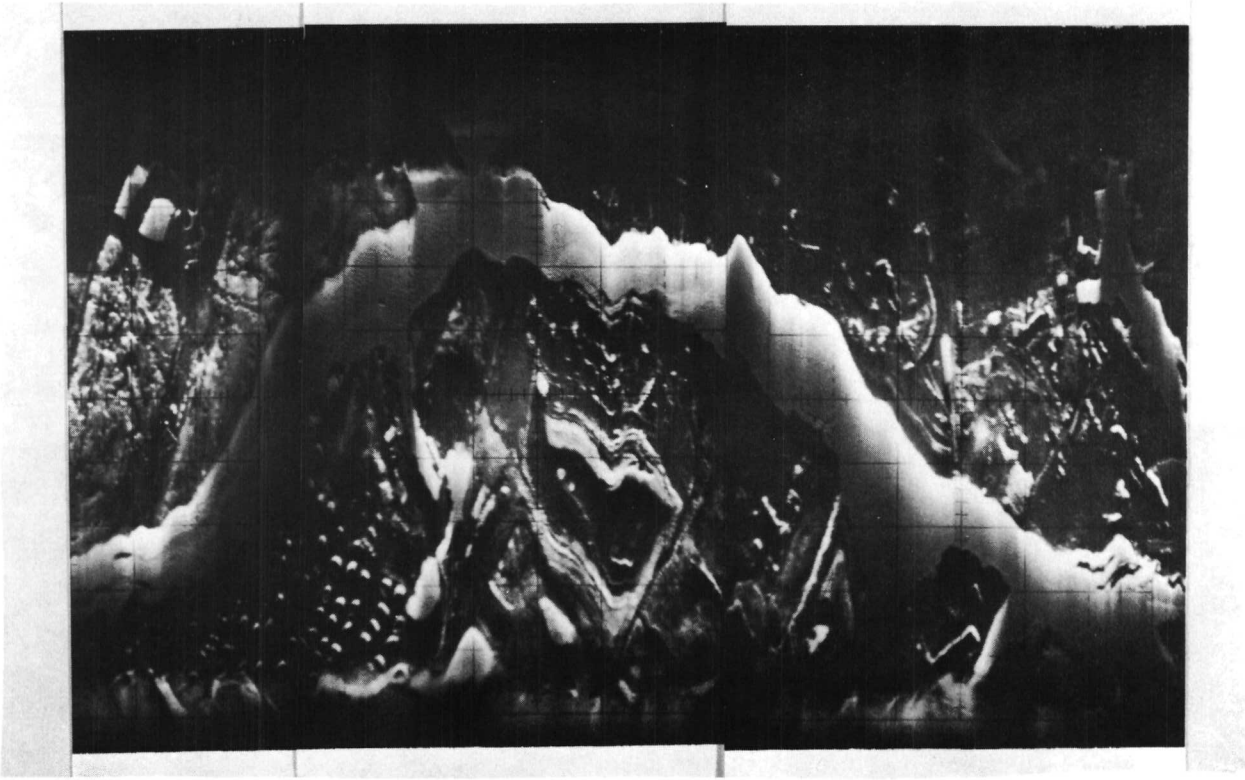


Fig. IV-6b Mosaic of East Half of Flight Line 8

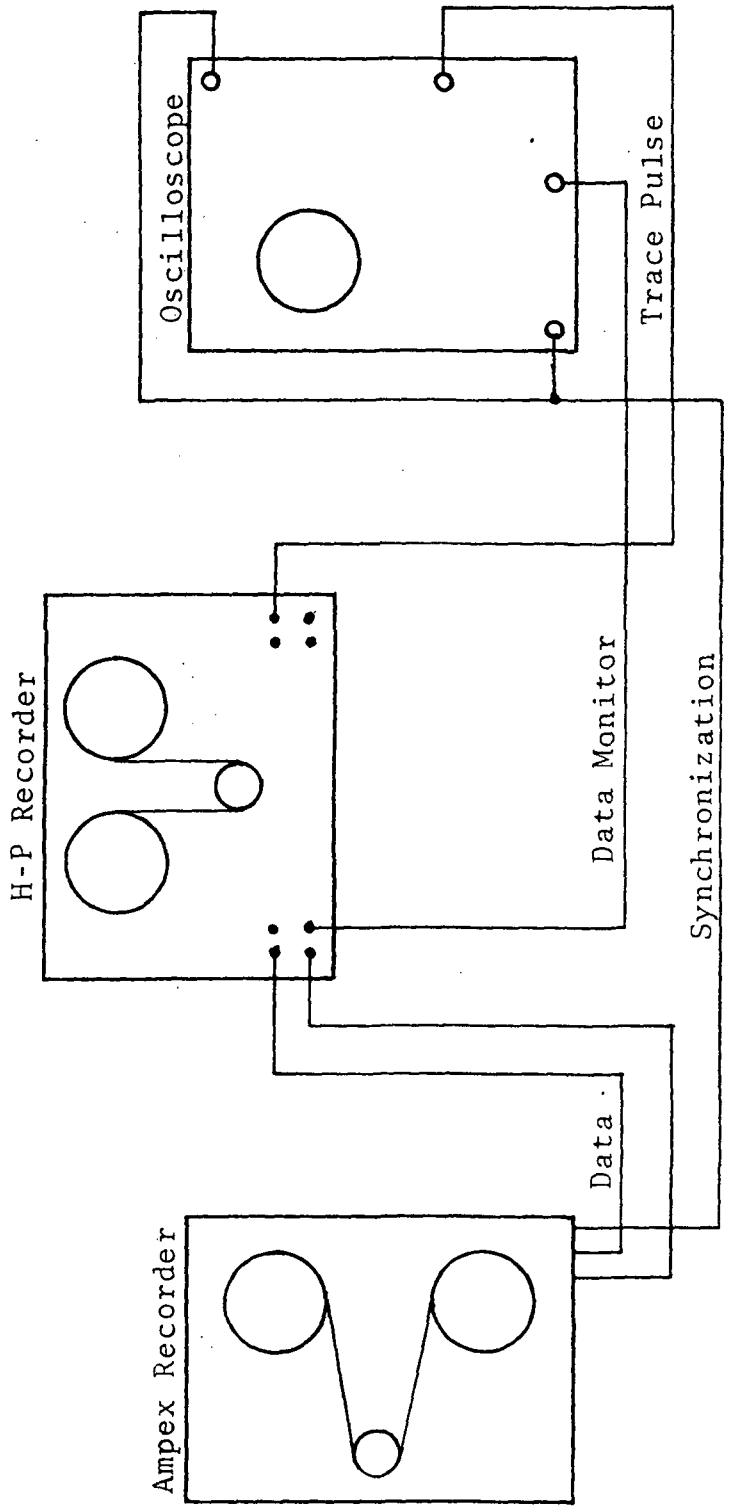


Fig. IV-7 Data Reproduction System

mation were reproduced. Table IV-9 lists the data channels and explains why each was or was not reproduced.

After the data was digitized, it had a resolution of approximatley 100 x 100 feet. While trying to use the data, it was found that the digital tape for flight line 8 and part of the digital tape for data channels 4 and 13, flight line 7 were defective and could not be read. The remaining digitized scanner data covered fixes 44.5 through 51.0 excluding 46.5, 47.0, 47.5, 48.5, and 49.0.

TABLE IV-9

LISTING OF REPRODUCED AND NONREPRODUCED DATA CHANNELS

CHANNEL	REPRODUCED	REMARKS
1	No	Overlapped channel 2
2	Yes	Channel 2 and 6 expected to have most information
3	No	Overlapped channel 4
4	Yes	Channel 4 and 13 were well separated from channel 2 and 6
5	No	Overlapped channel 4 and 6
6	Yes	Channel 2 and 6 expected to have most information
7	No	Synchronization
8	No	Data was defective
9	No	Recorded sequentially - would have required substantially more processing than other channels
10		
11		
12	No	Outside the range of interest of thesis
13	Yes	Channel 4 and 13 were well separated from channel 2 and 6
14	No	Synchronization

CHAPTER V

RESULTS

Model Selection

Theory

As an illustration of the concepts discussed in chapter III, the methods were applied to the data described in chapter IV. The results of this analysis are presented in this chapter.

The first step in applying the methods was to consider models suggested by the background of chapter II, that is, to use the equations developed in chapter II to relate waste concentrations to the quantitative data provided by the sensors. Scanner data consists of relative radiances, therefore it requires an expression relating waste concentration and radiance. Such an expression is provided by (II-23), restated below as (V-1).

$$N_a = Y \frac{1 + K'W}{a + bW} \exp(-A \sec \theta_{A1}) \qquad \text{(V-1)}$$

An expression for waste concentration in terms of radiance is obtained by solving (V-1) for W to obtain

$$W = \frac{-(Y \exp(-A \sec \theta_{A1}) - aN_a)}{(YK' \exp(-A \sec \theta_{A1}) - bN_a)} \qquad \text{(V-2a)}$$

$$W = C_0\left(\frac{C_1 - N_a}{C_2 - N_a}\right) \qquad \text{(V-2b)}$$

where: $C_0 = -a/b$

$C_1 = Y \exp(-A \sec\theta_{A1})/a$

$C_2 = -K' C_0 C_1$

A linear relationship between waste concentration and radiance can be obtained by expanding the second factor in the right member of (V-2b) in a Maclaurin's series. This gives

$$W = C_0\left(\frac{C_1}{C_2} + \frac{C_1 - C_2}{C_2^2} N_a + \ldots + \frac{C_1 - C_2}{C_2^{n+1}} N_a^n + \ldots\right) \qquad \text{(V-3)}$$

which suggests the use of a mathematical model of the form,

$$P = K_0 + \sum_{i=1}^{p} \sum_{j=1}^{m_i} K_{ij} N_{ai}^j \qquad \text{(V-4)}$$

where: P = value of water quality parameter

K_0 = constant

K_{ij} = constant coefficient of jth term for ith variable

p = number of variables

m_i = number of terms included for ith variable

If this model fits the measurements data, the H.L.L. method can be used to determine the coefficients, and to determine

which K_{ij}'s should be zero.

Photographic data consists of values of transparency densities. Equation (II-38), given below as (V-5), relates radiance and density

$$N_a = \frac{\exp\left(\frac{D_o - M}{G}\right) \exp(A' \sec\theta_{A1})}{Y' T \cos^4 \phi_2} \qquad \text{(V-5)}$$

Equating (V-1) and (V-5), then solving for W gives

$$W = C_o \left(\frac{C_3 - \exp(D_o/G)}{C_4 - \exp(D_o/G)} \right) \qquad \text{(V-6)}$$

where: $C_3 = \frac{Y Y' T \cos^4 \phi_2 \exp(M G)}{a \exp(A' \sec\theta_{A2} + A \sec\theta_{A1})}$

$C_4 = -K' C_o C_3$

The first step in linearizing (V-6) is to define a new variable,

$$x_1 \equiv \exp(D_o/G) \qquad \text{(V-7a)}$$

$$x_1 = (\exp D_o)^{1/G} \qquad \text{(V-7b)}$$

where: x_1 = new independent variable

and substitute for $\exp(D_o/G)$ in (V-6) to obtain an equation of the form of (V-2b). This can be expanded in the Maclaurin's series as

$$W = C_o \left(\frac{C_3}{C_4} + \frac{C_3 - C_4}{C_4^2} x_1 + \ldots + \frac{C_3 - C_4}{C_4^{n+1}} x_1^n + \ldots \right) \qquad \text{(V-8)}$$

The value of x_1, cannot be calculated because G is unknown, but by defining

$$x_2 = \exp D_o \tag{V-9}$$

where: x_2 = new independent variable

and substituting for exp D_o in (V-7b),

$$x_1 = x_2^{1/G} \tag{V-10}$$

Combining (V-10) and (V-8) gives

$$W = C_o\left(\frac{C_3}{C_4} + \frac{C_3 - C_4}{C_4^2} x_2^{1/G} + \ldots + \frac{C_3 - C_4}{C_4^{n+1}} x_2^{n/G} + \ldots\right) \tag{V-11}$$

which also has the form of (V-4).

If the model of (V-4) does not fit the data, mathematical models based on the ratios of radiances in different bands, as suggested by James and Burgess [6], might be tried.

Scatter Diagrams

For the scanner data, an indication of the applicability of the model of (V-4) can be determined by preparing scatter diagrams of the data. Figures A-1 through A-12 (pp. 139-150) are plots of turbidity, transmittance, and color versus radiances in the different

bands. The limited number of data points made it impossible to test many independent variables, therefore, the only independent variables that were considered were the first powers of the radiances. The scatter diagrams showed that this was at least a reasonable model to consider for the scanner data.

Scatter diagrams for the photographic data are given in figures A-13 through A-21 (pp. 151-159). In accordance with (V-9), the values of the water quality parameters were plotted versus the exponential of the transparency densities in the different bands. Only the first powers of the exponentials were used because of the small number of data points. In fact, the number of data points was so small that it was impossible to draw any conclusions about the applicability of (V-4) as a model for the photography, but to demonstrate the analysis method, (V-4) was used with only first powers.

Regressions

The second step in applying the methods of chapter III was to perform the regression and optimal parameter selection. This was done by using the SELECT subroutines written by L. R. LaMotte to implement the H.L.L. method. The subroutines were used to give the optimal one, two,

and three variable regressions for the scanner data and the optimal one and two variable regressions for the photographic data.

Scanner Data

Tables V-1 through V-3 give the full regressions and the various optimal regressions from the scanner data for turbidity, transmittance, and color. These tables also give the t-value, variance, confidence level, and the deviation for the .9 confidence interval for the coefficients. Tables V-4 through V-6 give the analysis of variance plus the square of the correlation coefficient, actual F-value, and F-value required for a .9 confidence level for each regression.

Photographic Data

Results for the photographic data are given in tables V-7 through V-9, which give the regressions, and in tables V-10 through V-12, which give the analysis of variance for each regression. It should be noted that the number of data points is four, just one more than the number of variables in the full regression, hence the regression will fit the data perfectly. This makes tests on the full regression impossible and it is only included in the tables as a matter of interest.

TABLE V-1

REGRESSIONS FOR TURBIDITY (SCANNER DATA)

Regression	Variable	Coeff.	t-value	Variance	Conf.Level	Deviation for .9 Conf.Inter.
Full	Ch. 2	-66.98	-1.74	1483.	.84	± 90.61
	Ch. 4	-79.93	-1.11	5217.	.67	±169.95
	Ch. 6	-23.63	- .14	29830.	.10	±406.40
	Ch. 13	-37.26	- .88	1789.	.57	± 99.52
	Constant	276.88				
Three Variable	Ch. 2	-67.58	-1.97	1176.	.89	± 73.11
	Ch. 4	-84.89	-1.52	3137.	.81	±119.41
	Ch. 13	-35.97	- .97	1365.	.63	± 78.77
	Constant	255.24				
Two Variable	Ch. 2	-64.26	-1.89	1154.	.89	± 68.45
	Ch. 4	-111.09	-2.27	2391.	.94	± 98.53
	Constant	238.84				
One Variable	Ch. 13	-37.85	-1.28	880.	.76	± 57.64
	Constant	76.35				

TABLE V-2

REGRESSION FOR TRANSMITTANCE (SCANNER DATA)

Regression	Variable	Coeff.	t-value	Variance	Conf.Level	Deviation for .9 Conf.Inter.
Full	Ch. 2	-3.39	-.62	29.27	.44	±12.73
	Ch. 4	-3.94	-.38	102.99	.28	±23.87
	Ch. 6	41.13	1.69	588.89	.84	±57.10
	Ch. 13	10.73	1.80	35.31	.86	±13.98
	Constant	42.32				
Three Variable	Ch. 2	-2.13	-.54	15.60	.39	± 8.42
	Ch. 6	36.39	1.90	365.70	.88	±40.77
	Ch. 13	9.53	2.05	21.45	.91	± 9.87
	Constant	42.96				
Two Variable	Ch. 6	39.17	2.26	299.22	.94	±34.85
	Ch. 13	10.92	3.02	13.09	.98	± 7.29
	Constant	35.56				
One Variable	Ch. 13	12.94	2.92	19.55	.98	± 8.59
	Constant	77.23				

TABLE V-3

REGRESSIONS FOR COLOR (SCANNER DATA)

Regression	Variable	Coeff.	t-value	Variance	Conf.Level	Deviation for .9 Conf.Inter.
Full	Ch. 2	- 7.31	- .100	5315.	.08	±171.54
	Ch. 4	-37.49	- .274	18703.	.20	±321.79
	Ch. 6	-186.40	- .570	106942.	.40	±769.48
	Ch. 13	-105.50	-1.317	6412.	.74	±188.42
	Constant	479.48				
Three Variable	Ch. 4	-29.28	- .298	9632.	.22	±209.24
	Ch. 6	-190.13	- .653	84656.	.46	±620.32
	Ch. 13	-104.92	-1.467	5117.	.80	±152.51
	Constant	464.53				
Two Variable	Ch. 6	-237.43	-1.057	50479.	.67	±452.72
	Ch. 13	-119.88	-2.551	2208.	.96	± 94.68
	Constant	501.62				
One Variable	Ch. 13	-132.09	-2.876	2109.	.98	± 89.23
	Constant	249.09				

TABLE V-4

ANALYSIS OF VARIANCE FOR TURBIDITY (SCANNER DATA)

Regression	Source	DF	SS	MS	R^2	F-value	.9 F-value
Full	Total	8	142.72				
	Regression	4	78.76	19.69	.55	1.23	5.34
	Error	4	63.96	15.99			
Three Variable	Total	8	142.72				
	Regression	3	78.46	26.15	.55	2.03	4.19
	Error	5	64.26	12.85			
Two Variable	Total	8	142.72				
	Regression	2	66.28	33.14	.46	2.60	3.78
	Error	6	76.44	12.74			
One Variable	Total	8	142.72				
	Regression	1	26.94	26.94	.19	1.63	3.78
	Error	7	115.78	16.54			

TABLE V-5

ANALYSIS OF VARIANCE FOR TRANSMITTANCE (SCANNER DATA)

Regression	Source	DF	SS	MS	R^2	F-value	.9 F-value
Full	Total	8	5.722				
	Regression	4	4.460	1.114	.78	3.53	5.34
	Error	4	1.263	.316			
Three Variable	Total	8	5.722				
	Regression	3	4.412	1.471	.77	5.61	4.19
	Error	5	1.310	.262			
Two Variable	Total	8	5.722				
	Regression	2	4.335	2.168	.76	9.38	3.78
	Error	6	1.387	.231			
One Variable	Total	8	5.722				
	Regression	1	3.150	3.150	.55	8.57	3.78
	Error	7	2.573	.368			

TABLE V-6

ANALYSIS OF VARIANCE FOR COLOR (SCANNER DATA)

Regression	Source	DF	SS	MS	R^2	F-value	.9 F-value
Full	Total	8	605.6				
	Regression	4	376.2	94.06	.62	1.64	5.34
	Error	4	229.3	57.33			
Three Variable	Total	8	605.6				
	Regression	3	375.7	125.22	.62	2.72	4.19
	Error	5	229.9	45.98			
Two Variable	Total	8	605.6				
	Regression	2	371.6	185.79	.61	4.76	3.78
	Error	6	234.0	39.00			
One Variable	Total	8	605.6				
	Regression	1	328.0	328.0	.54	8.27	3.78
	Error	7	277.5	39.6			

TABLE V-7

REGRESSIONS FOR TURBIDITY (PHOTOGRAPHIC DATA)

Regression	Variable	Coeff.	t-value	Variance	Conf.Level	Deviation for .9 Conf.Inter.
Full	Red	260.12				
	Green	-473.95				
	Blue	102.45				
	Constant	307.69				
Two Variable	Red	189.50	1.84	10609.	.80	± 650.34
	Green	-231.36	-1.34	29886.	.72	±1091.54
	Constant	81.54				
One Variable	Red	67.95	1.18	3294.	.65	± 167.59
	Constant	-202.83				

TABLE V-8

REGRESSIONS FOR TRANSMITTANCE (PHOTOGRAPHIC DATA)

Regression	Variable	Coeff.	t-value	Variance	Conf.Level	Deviation for .9 Conf.Inter.
Full	Red	-32.26				
	Green	67.66				
	Blue	-23.54				
	Constant	59.54				
Two Variable	Red	- 7.23	- .542	177.90	.41	±84.22
	Blue	- 4.59	- .340	182.38	.28	±85.27
	Constant	129.38				
One Variable	Red	- 9.78	-1.186	67.94	.65	±24.07
	Constant	126.15				

TABLE V-9

REGRESSIONS FOR COLOR (PHOTOGRAPHIC DATA)

Regression	Variable	Coeff.	t-value	Variance	Conf.Level	Deviation for .9 Conf.Inter.
Full	Red	295.55				
	Green	-379.99				
	Blue	11.45				
	Constant	200.11				
Two Variable	Red	287.65	24.97	132.69	.99	± 72.73
	Green	-352.86	-18.25	373.78	.99	±122.07
	Constant	174.82				
One Variable	Red	102.27	1.46	4932.26	.72	±205.07
	Constant	-258.90				

TABLE V-10

ANALYSIS OF VARIANCE FOR TUBIDITY (PHOTOGRAPHIC DATA)

Regression	Source	DF	SS	MS	R^2	F-value	.9 F-value
Two Variable	Total	3	372.19				
	Regression	2	293.79	146.89	.79	1.87	49.50
	Error	1	78.40	78.40			
One Variable	Total	3	372.19				
	Regression	1	153.36	153.36	.41	1.40	8.53
	Error	2	218.82	109.41			

TABLE V-11

ANALYSIS OF VARIANCE FOR TRANSMITTANCE (PHOTOGRAPHIC DATA)

Regression	Source	DF	SS	MS	R^2	F-value	.9 F-value
Two Variable	Total	3	7.688				
	Regression	2	3.643	1.821	.47	.45	49.50
	Error	1	4.045	4.045			
One Variable	Total	3	7.688				
	Regression	1	3.175	3.175	.41	1.41	8.53
	Error	2	4.513	2.256			

TABLE V-12

ANALYSIS OF VARIANCE FOR COLOR (PHOTOGRAPHIC DATA)

Regression	Source	DF	SS	MS	R^2	F-value	.9 F-value
Two Variable	Total	3	675.00				
	Regression	2	674.02	337.01	1.00	343.70	49.50
	Error	1	.98	.98			
One Variable	Total	3	675.00				
	Regression	1	347.37	347.37	.51	2.12	8.53
	Error	2	327.63	163.81			

Conclusions

It is doubtful that any meaningful conclusions can be drawn from the data due to the limited quantity and variation that were available. The data analysis, however, does illustrate the application of the analysis method and does show the use of the method in relating water quality parameters and spectral signatures.

CHAPTER VI

CONCLUSIONS AND RECOMMENDATIONS

Review

The primary aim of this study was to apply regression analysis to the problem of determining water quality parameters from spectral signatures. This was accomplished in two phases. The first phase, discussed in chapter III, was concerned with the equations required to perform a regression, with tests of the strength and reliability of a regression, and with a method of selecting optimal subsets of a set of independent parameters. The second phase was concerned with determining suitable mathematical models relating water quality parameters to the data available from airborne multispectral scanners and photography. The background for the second phase was provided by chapter II, and the actual model selection is described by the first part of chapter V.

A secondary aim of this study was to apply regression analysis techniques to a particular set of remote sensing data. This involved: (1) modifying the models to obtain the best possible fit to the data, (2) performing the regression analysis, (3) testing the strength and reli-

ability of the regressions, (4) selecting the optimal bands, from those available, for measuring water quality parameters, and (5) checking that the predictive capability of the regressions was sufficiently accurate to measure water quality parameters. Efforts to satisfy the secondary aim are described in chapter V.

The particular data set to which the regression analysis was applied was obtained along the Houston Ship Channel. The ground observations of turbidity, transmittance, and color were performed approximately the same time that the multispectral scanner data and photography were collected by an aircraft overflight.

Conclusions

In chapter III are presented methods for performing a regression, tests for the strength and reliability of the regression, and an algorithm for reducing the computational requirements of selecting an optimal subset of independent variables. Chapter II provides the background for proposing a mathematical model and the first part of chapter V describes the implementation of model selection.

The secondary aim of this study required a realistic set of data. After data were obtained from the Houston Ship Channel, difficiencies in the data and processing

difficulties were encountered. First, it was found that the water quality parameters were essentially uniform along the ship channel, and therefore the range of data values was limited. Second, the NASA facilities that were to digitize the scanner data had been modified and could no longer process the measurements. This required that the data be digitized by other means which resulted in some degradation of the data. Third much of the scanner data was useless because it was impossible to obtain synchronization on one of the scanner tapes, and because parts of the digital tapes containing the digitized scanner data were defective and were impossible to read.

The limited range of data values had two adverse effects: (1) the small range of data values made random variations in the data relatively more significant, and regression analysis is not designed to account for random variations, (2) a regression is only applicable over the range of values included in the data used to obtain the regression, hence, the limited range of data values limited the range of applicability of the regressions.

The limited number of data samples also had two adverse effects, it made it difficult to differentiate between the fit of different models and it placed more stringent requirements on the regression tests. Fitting

a model becomes difficult because there are too few points to make concentrations of points stand out. More stringent requirements on the tests come about because a regression must fit a few data points more closely than it must fit many points in order to achieve the same confidence level.

Although regressions were performed and tested on the data that were available, the limitations imposed by the data made the results less meaningful than had been hoped.

Recommendations

A better indication of the applicability of regression analysis for measuring water quality parameters with spectral signatures would be obtained by applying regression analysis to a data set with a much larger number of data samples and a wider range of values.

An application for regression analysis might be ERTS-1 data which consists of four bands of multispectral scanner data. Figure VI-1 shows some recently acquired imagery from the Great Salt Lake in Utah. Figure VI-2 shows a gray scale computer printout of this region. Variations in the appearance of the water as recorded in the imagery are obvious. Table VI-1 gives the average

Fig. VI-1 ERTS-1 Imagery of Great Salt Lake

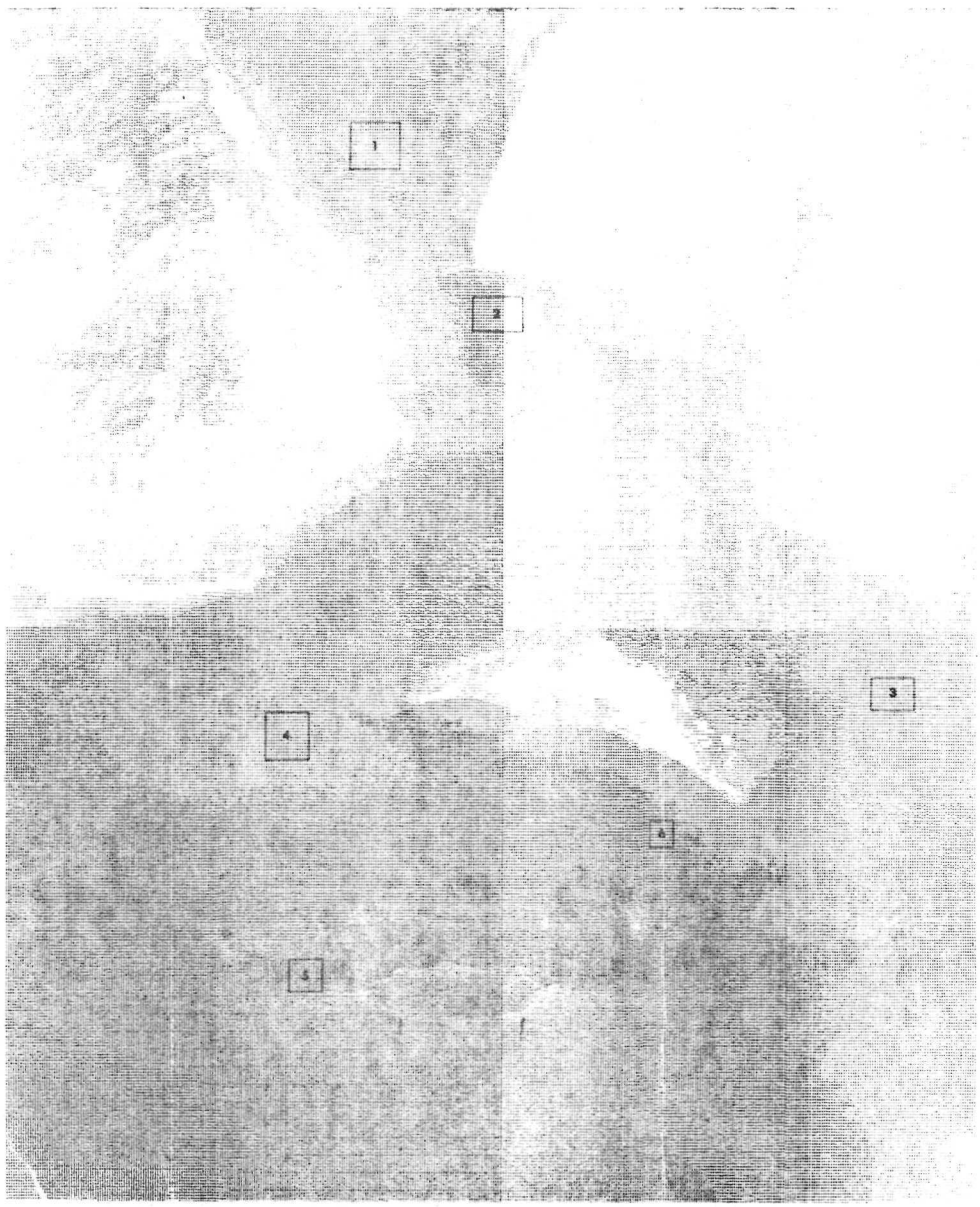


Fig. VI-2 Gray Scale Computer Printout of ERTS-1 Data from Great Salt Lake

TABLE VI-1

SIGNATURES* FROM ERTS-1 DATA

Region	Band (Microns)			
	.5-.6	.6-.7	.7-.8	.8-1.1
1	7.08	3.87	1.43	0.29
2	6.36	3.25	1.21	0.23
3	7.44	4.08	1.55	0.33
4	5.78	2.29	0.82	0.17
5	4.09	1.50	0.59	0.17
6	4.94	1.85	0.67	0.16

* The signatures are radiance in mw/cm^2-STR-μ

radiance value for each of the four bands for each of the six regions marked in figure VI-2 (p. 132). Differences in the signatures for the six regions are apparent. These data were only recently received and no attempt has been made to implement the regression analysis.

Water quality parameters are defined to give some indication of the condition of the water. The water quality parameters considered in this study were limited to turbidity, transmittance, and color because of the several standard water quality measurements, these transmission characteristics of the water were expected to give the best correlation with spectral signatures. The spectral signatures used in remote sensing, however, are reflective. This suggests that if remote sensing techniques are to be used to determine water quality, it might be well to define new water quality parameters based on the reflective characteristics of the water.

REFERENCES

[1] R. N. Colwell, "Uses and Limitations of Multispectral Remote Sensing," Proceedings of the Fourth International Symposium on Remote Sensing of the Environment, University of Michigan, Ann Arbor, April 1966.

[2] J. E. Estes and B. Golomb, "Monitoring Environmental Pollution," Journal of Remote Sensing, March-April 1966.

[3] G. C. Ewing, "Remote Spectrography of Ocean Color as an Index of Biological Activity," Proceedings of the Symposium on Remote Sensing in Marine Biology and Fishery Resources, Texas A&M University, College Station, Texas, March 1971.

[4] F. B. Silvestro, "Quantitative Remote Sensing of Water Pollution," Presented at 15th Annual Technical Meeting of the Institute of Environmental Sciences, Anaheim, California, April 1969.

[5] C. S. Yentsch, "The Influence of Phytoplankton Pigments on the Colour of Sea Water," Deep-Sea Research, Vol. 7, Pergamon Press Ltd., London, 1960.

[6] W. P. James and F. J. Burgess, "Pulp Mill Outfall Analysis by Remote Sensing Techniques," Tappi, The Journal of the Technical Association of the Pulp and Paper Industry, Vol. 54, No. 3, March 1971.

[7] D. S. Lowe, "Line Scan Devices and Why Use Them," Proceedings of the Fifth Symposium on Remote Sensing of the Environment, University of Michigan, Ann Arbor, April 1968.

[8] M. A. Bramson, Translated by R. B. Rodman, Infrared Radiation: A Handbook for Applications, Plenum Press, New York, 1968.

[9] L. F. Johnson, "On the Performance of Infrared Sensors in Earth Observations," Technical Report RSC-37, Remote Sensing Center, Texas A&M University, College Station, Texas, September 1971.

[10] W. P. James, "Quantitative Evaluation of Water Quality in the Coastal Zone by Remote Sensing," Technical Report RSC-33, Remote Sensing Center, Texas A&M University, College Station, Texas, September 1971.

[11] A. E. Conrady, C. R. Davidson, C. R. Gibson, W. B. Hislop, F. C. V. Laws, J. H. G. Monypenny, H. Moss, A. S. Newman, G. H. Rodman, S. E. Sheppard, W. L. F. Wastell, W. M. Webb, and H. S. L. Winterbotham, Photography as a Scientific Implement, D. Van Nostrand Company, New York, 1923.

[12] A. C. Conrod and K. A. Rottweiler, "Water Quality Measurements with Airborne Multispectral Scanners," Joint Conference on Sensing of Environmental Pollutants, Palo Alto, California, November 1971.

[13] R. A. Holmes and R. B. MacDonald, "The Physical Basis of System Design for Remote Sensing in Agriculture," Proceedings of IEEE, Vol. 57, No. 4, April 1969.

[14] G. Kanecny and E. E. Derenyi, "Geometrical Considerations for Mapping from Scan Imagery," Proceedings on Remote Sensing of the Environment, University of Michigan, Ann Arbor, April 1966.

[15] P. G. Hasell, Jr., Personal Correspondence, October 1971.

[16] N. R. Draper and H. Smith, Applied Regression Analysis, John Wiley & Sons, Inc., New York, 1966.

[17] D. L. Harnett, Introduction to Statistical Methods, Addison-Wesley Publishing Company, Reading, Massachusetts, May 1971.

[18] R. R. Hocking and R. N. Leslie, "Selection of the Best Subset in Regression Analysis," Technometrics, Vol. 9, No. 4, November 1967.

[19] L. R. LaMotte and R. R. Hocking, "Computational Efficiency in Selection of Regression Variables," Technometrics, Vol. 12, No. 1, February 1970.

[20] A. Papoulis, Probability, Random Variables, and Stochastic Processes, McGraw-Hill Book Company, New York, 1965.

[21] CRC Standard Mathematical Tables, Fourteenth Edition, Edited by S. M. Selby, The Chemical Rubber Company, Cleveland, Ohio, 1965.

[22] Standard Methods for the Examination of Water and Waste Water Including Bottom Sediments and Sludges, Twelfth Edition, American Public Health Association, Inc., New York, New York, 1965.

[23] H. U. Sverdrop, M. W. Johnson, and R. H. Fleming, The Oceans, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1942,

APPENDIX A

SCATTER DIAGRAMS

The following figures are plots of water quality parameters versus the relative radiance for the scanner data and versus the exponential of the density for the photographic data. Table A-1 lists the band for each of the scanner channels that were plotted.

TABLE A-1

BANDS FOR PLOTTED SCANNER CHANNELS

Scanner Channel	Band (microns)
Channel 2	.46-.49
Channel 4	.52-.57
Channel 6	.58-.64
Channel 13	.63-.70

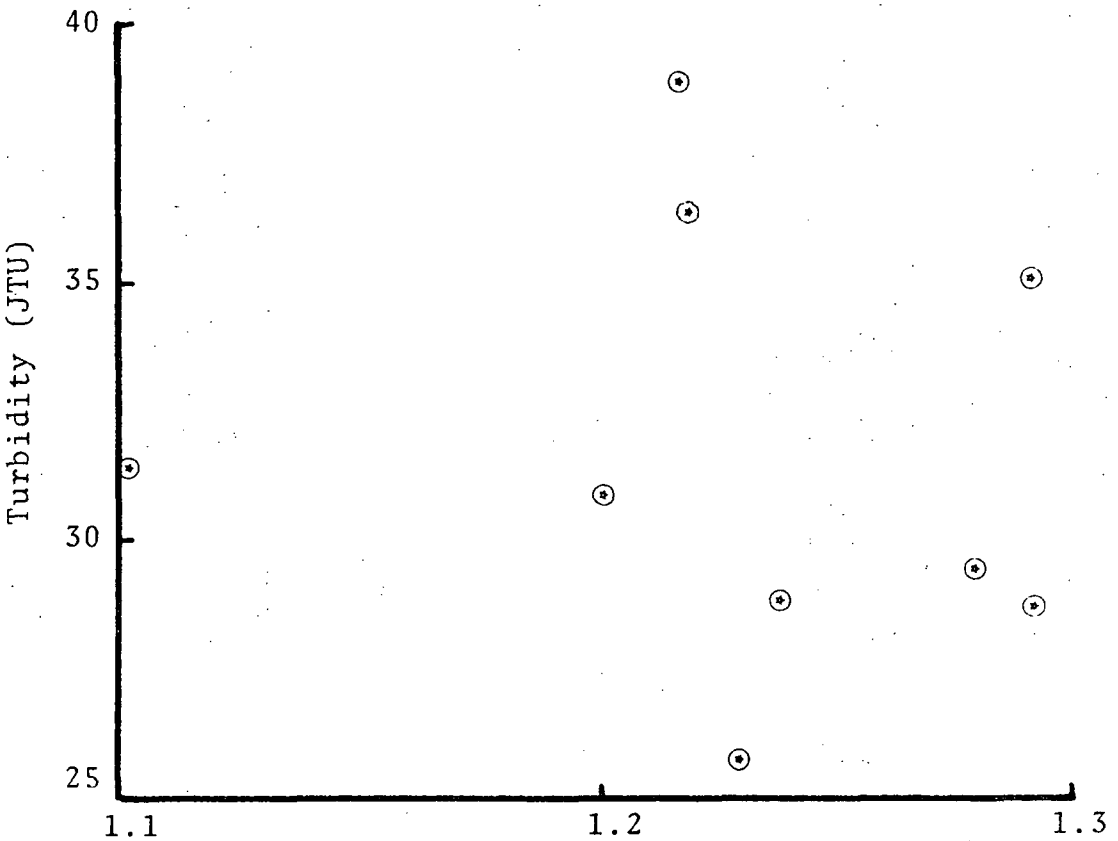


Fig. A-1 Scanner Data Scatter Diagram

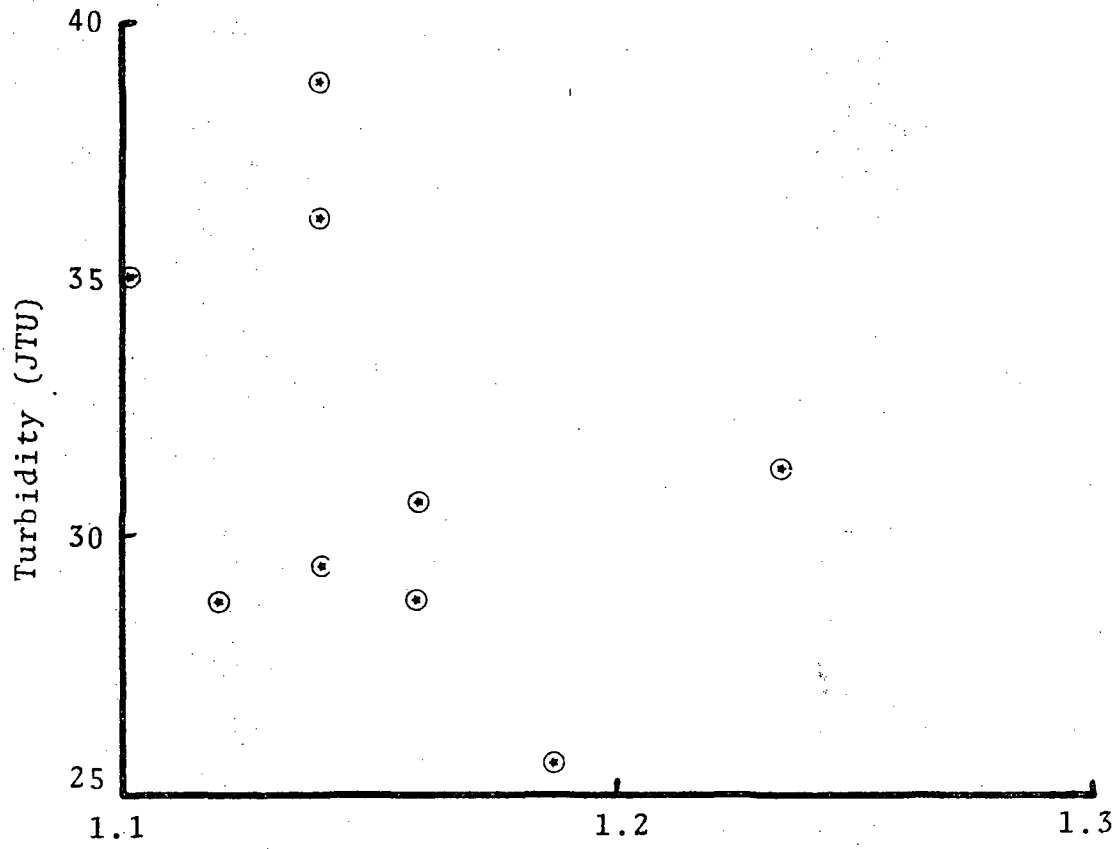


Fig. A-2 Scanner Data Scatter Diagram

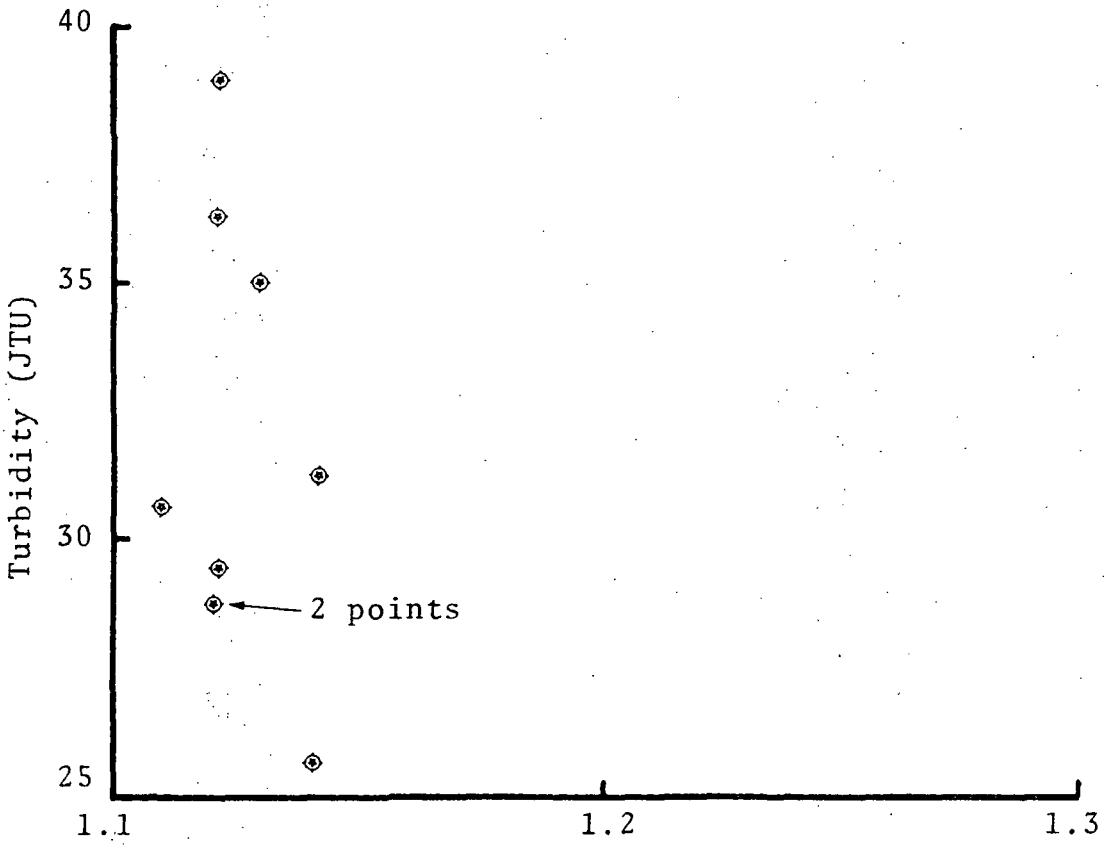


Fig. A-3 Scanner Data Scatter Diagram

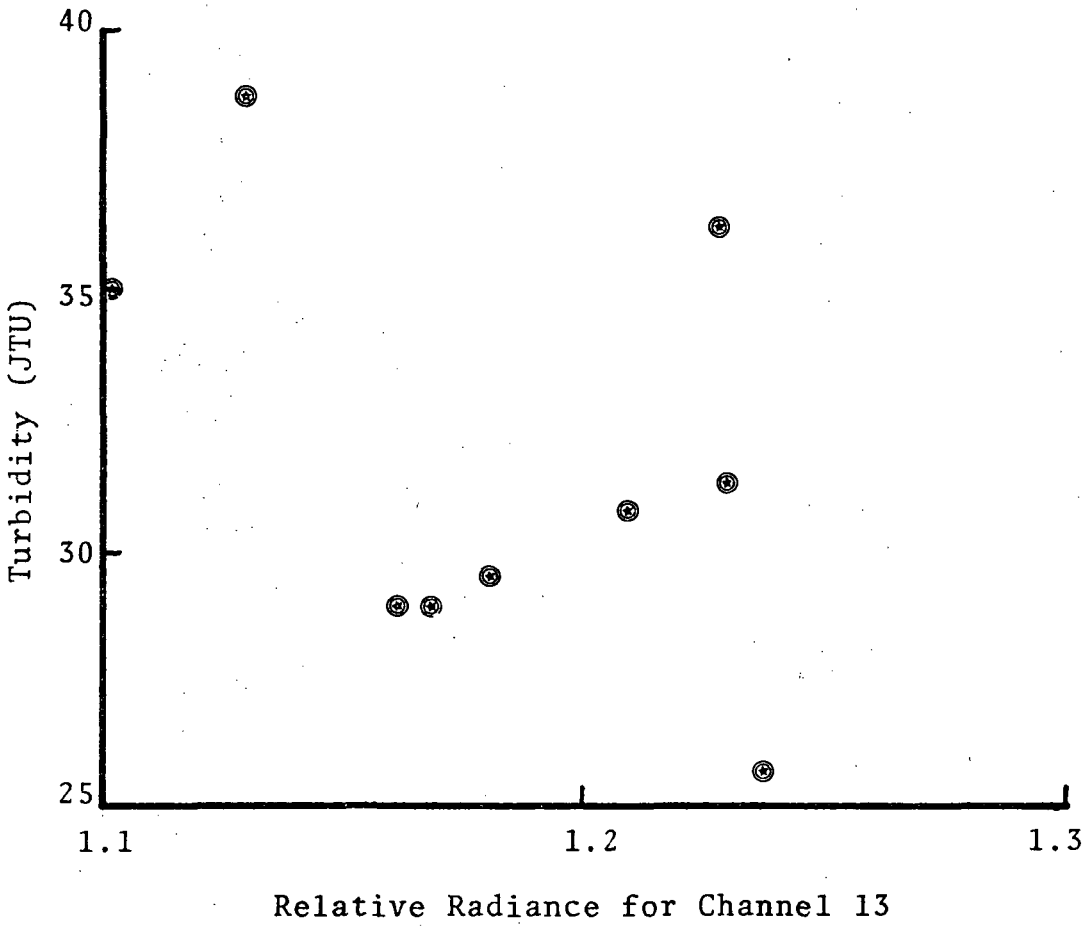


Fig. A-4 Scanner Data Scatter Diagram

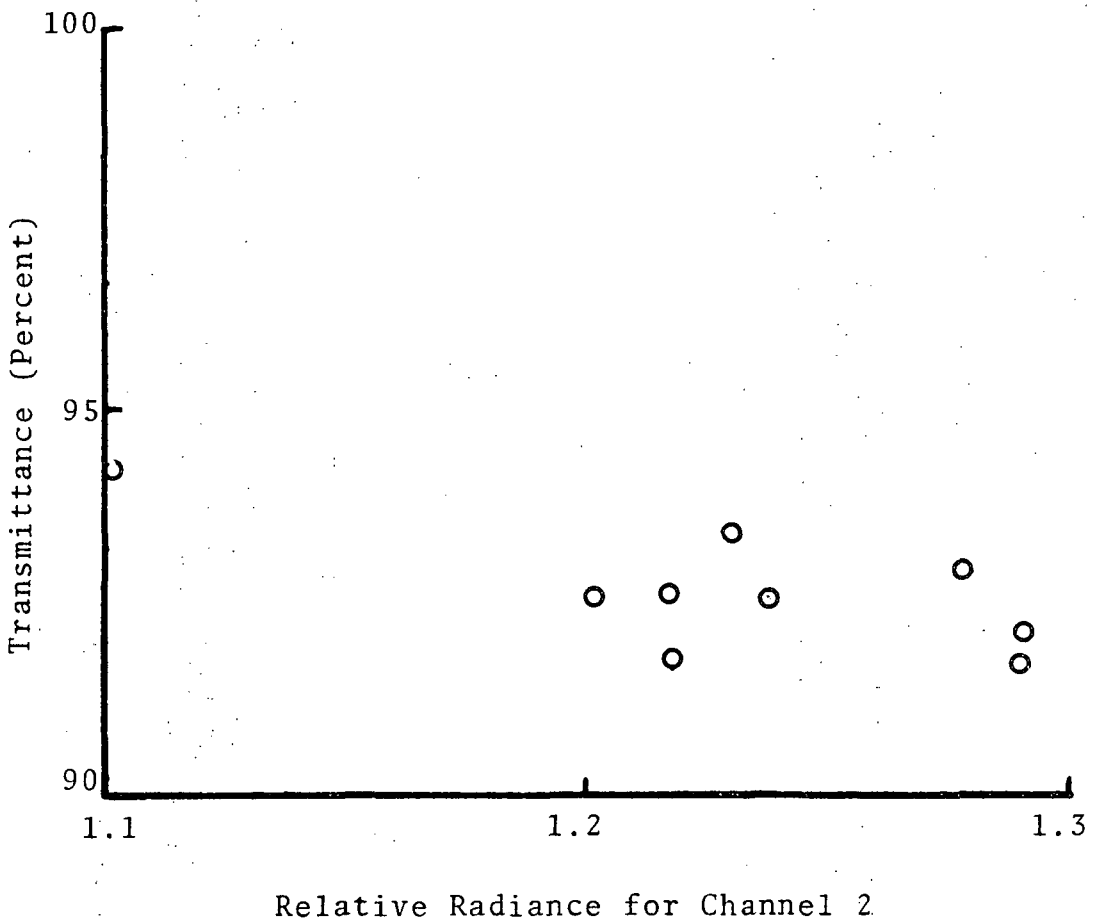


Fig. A-5 Scanner Data Scatter Diagram

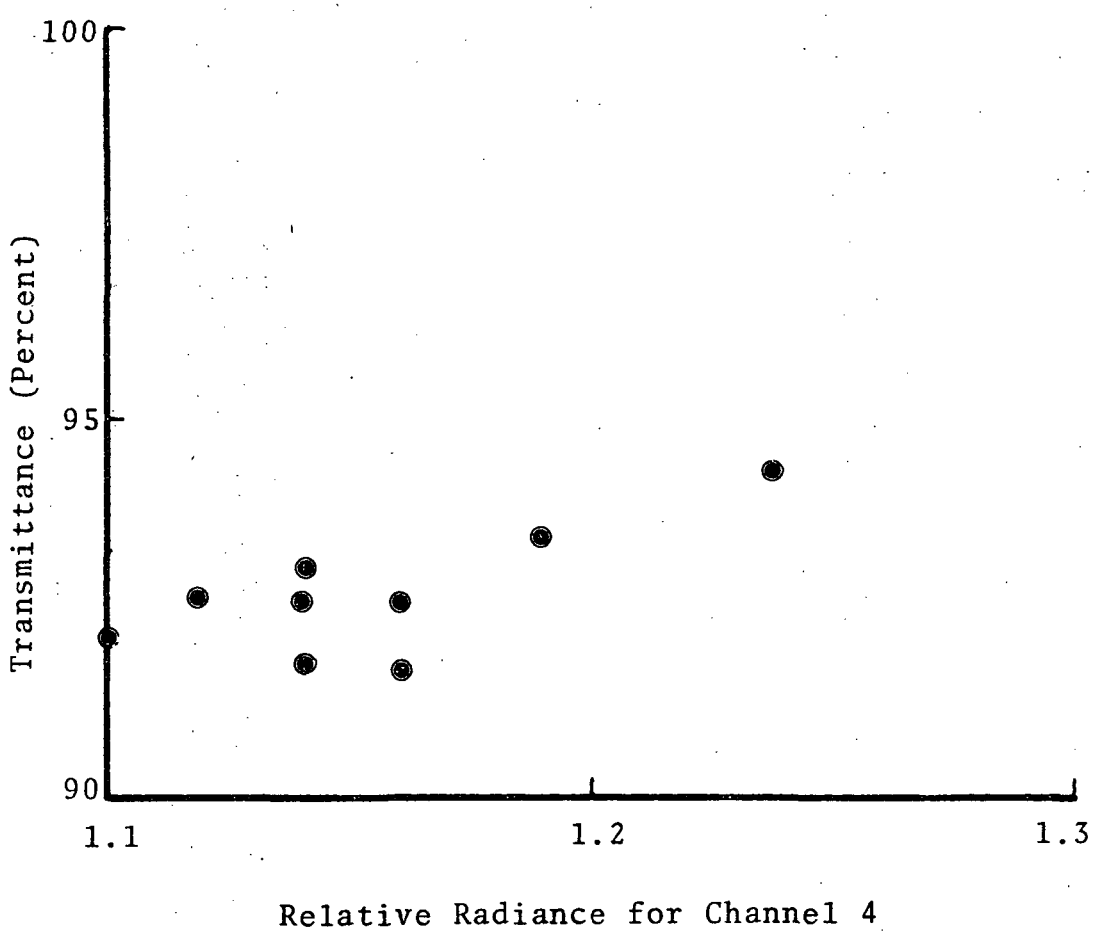


Fig. A-6 Scanner Data Scatter Diagram

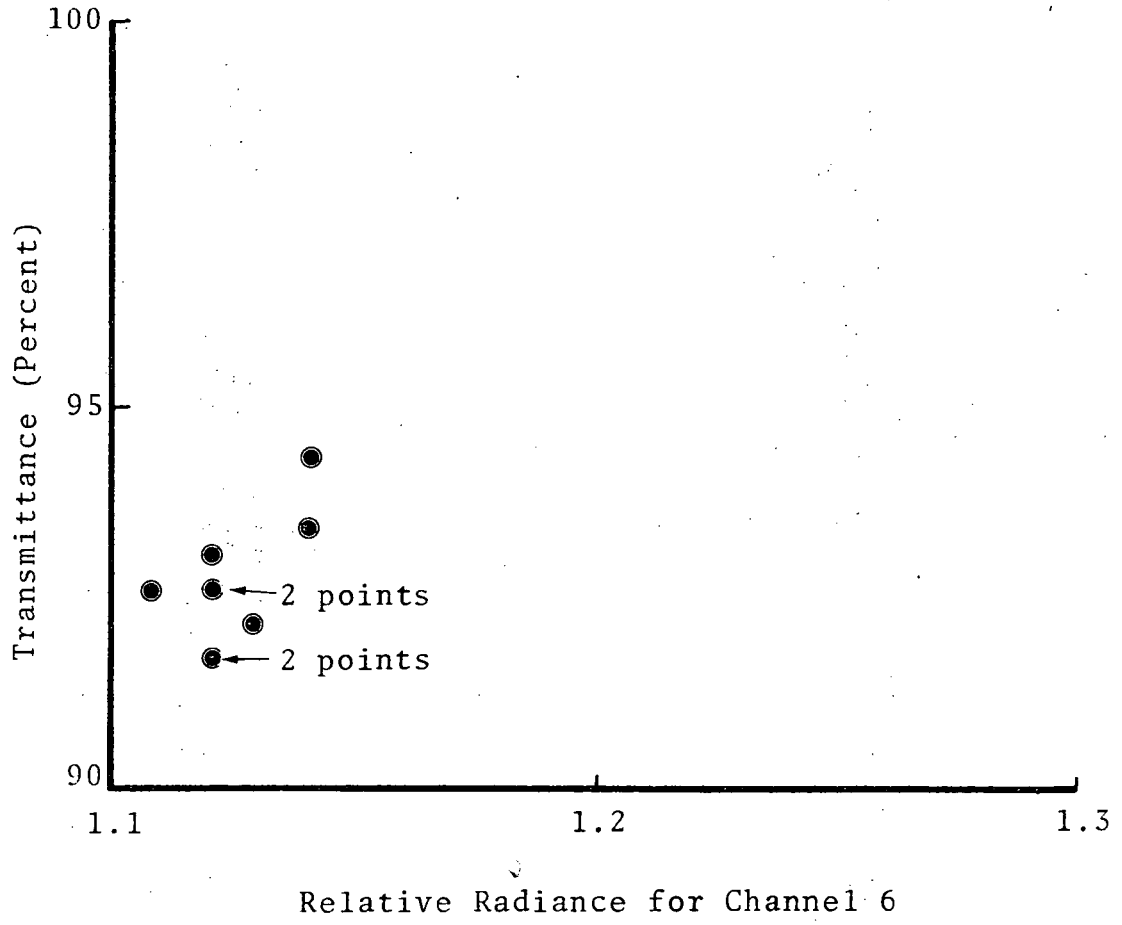


Fig. A-7 Scanner Data Scatter Diagram

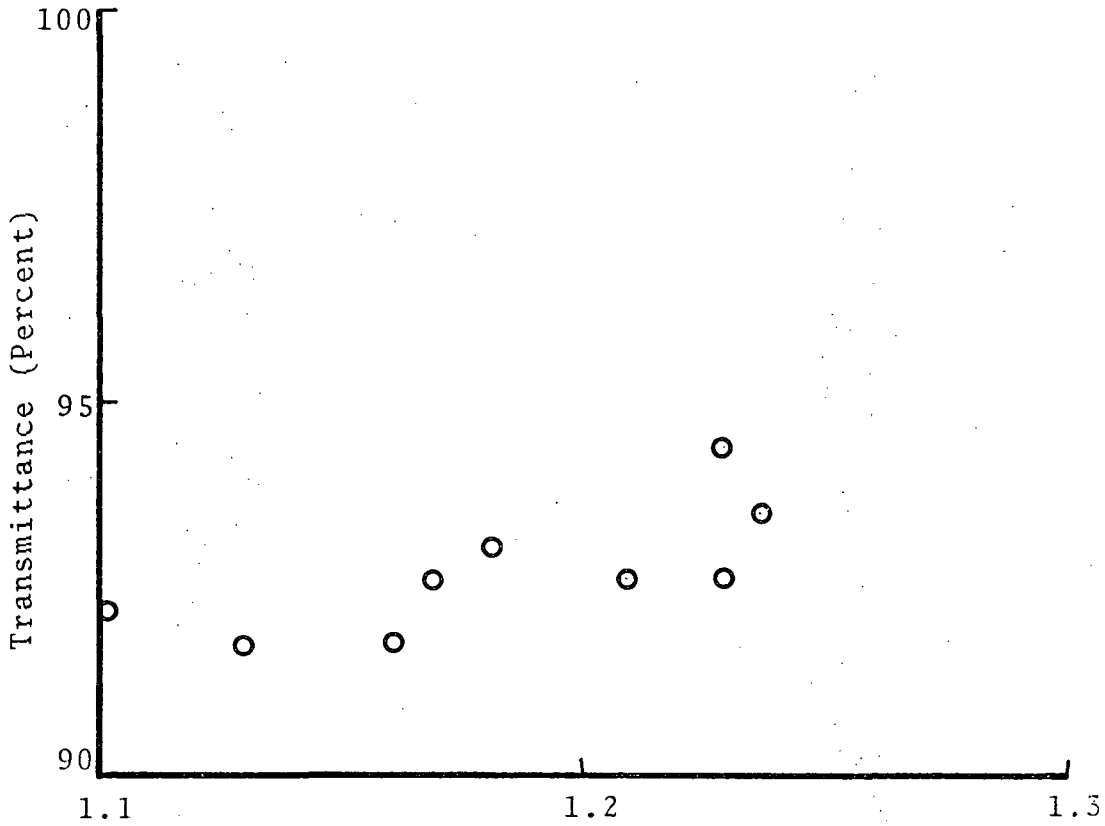


Fig. A-8 Scanner Data Scatter Diagram

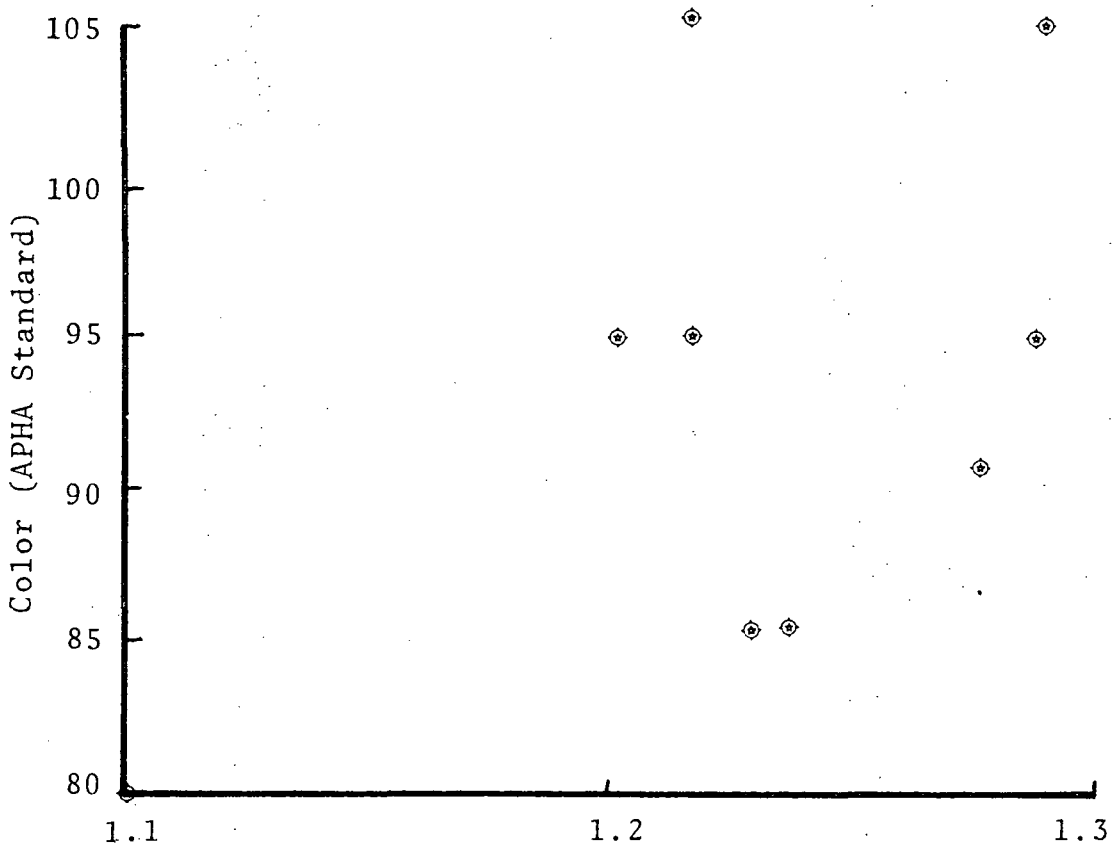


Fig. A-9 Scanner Data Scatter Diagram

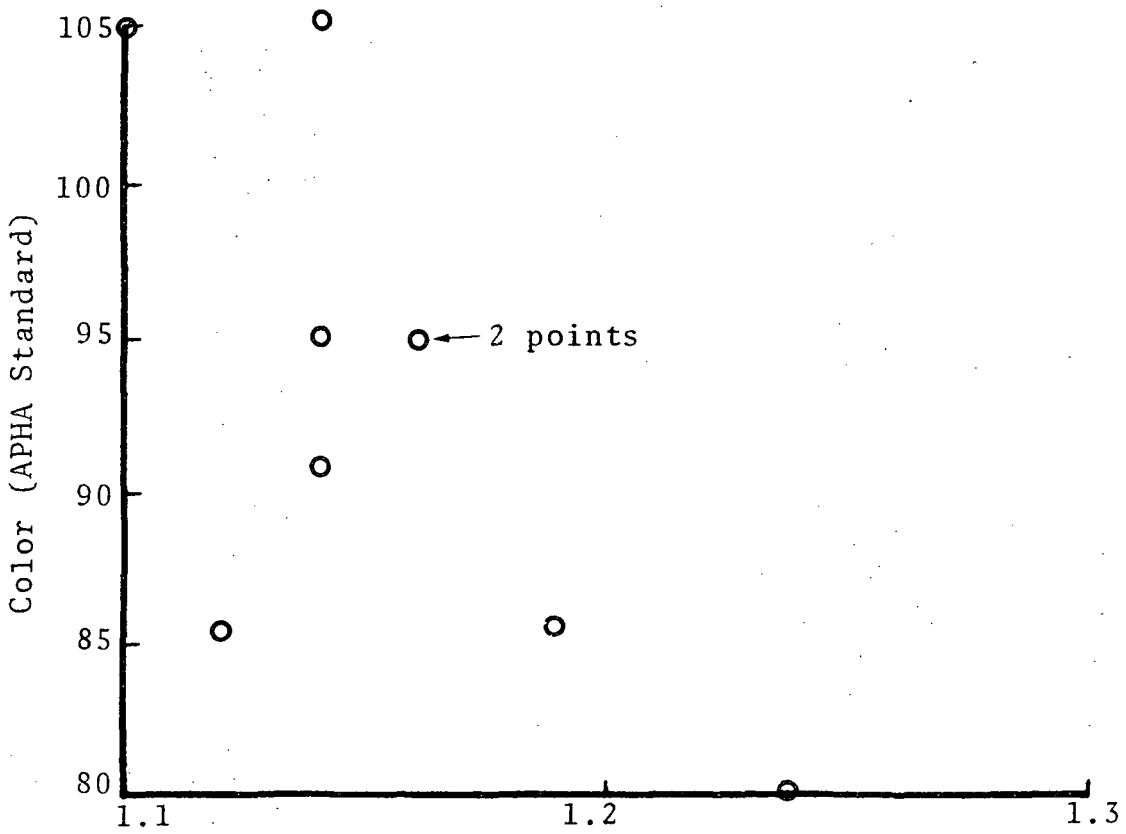


Fig. A-10 Scanner Data Scatter Diagram

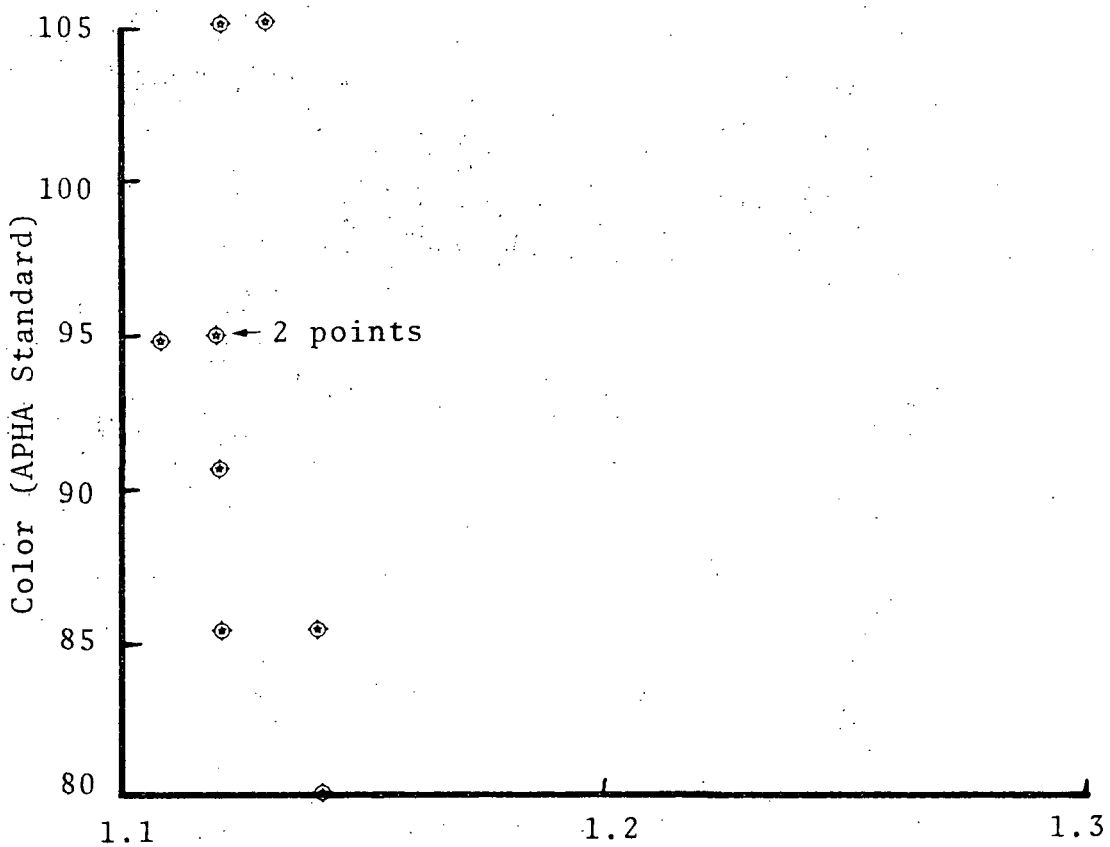


Fig. A-11 Scanner Data Scatter Diagram

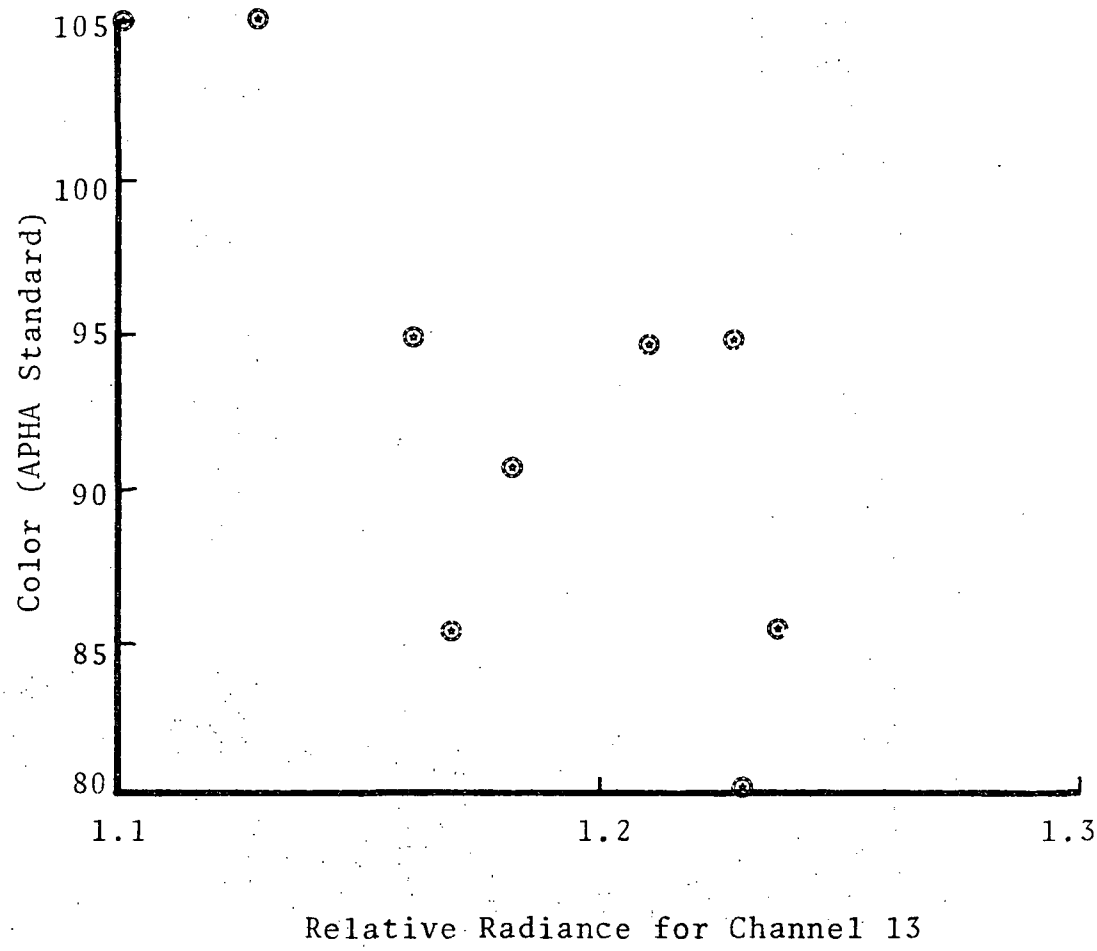


Fig. A-12 Scanner Data Scatter Diagram

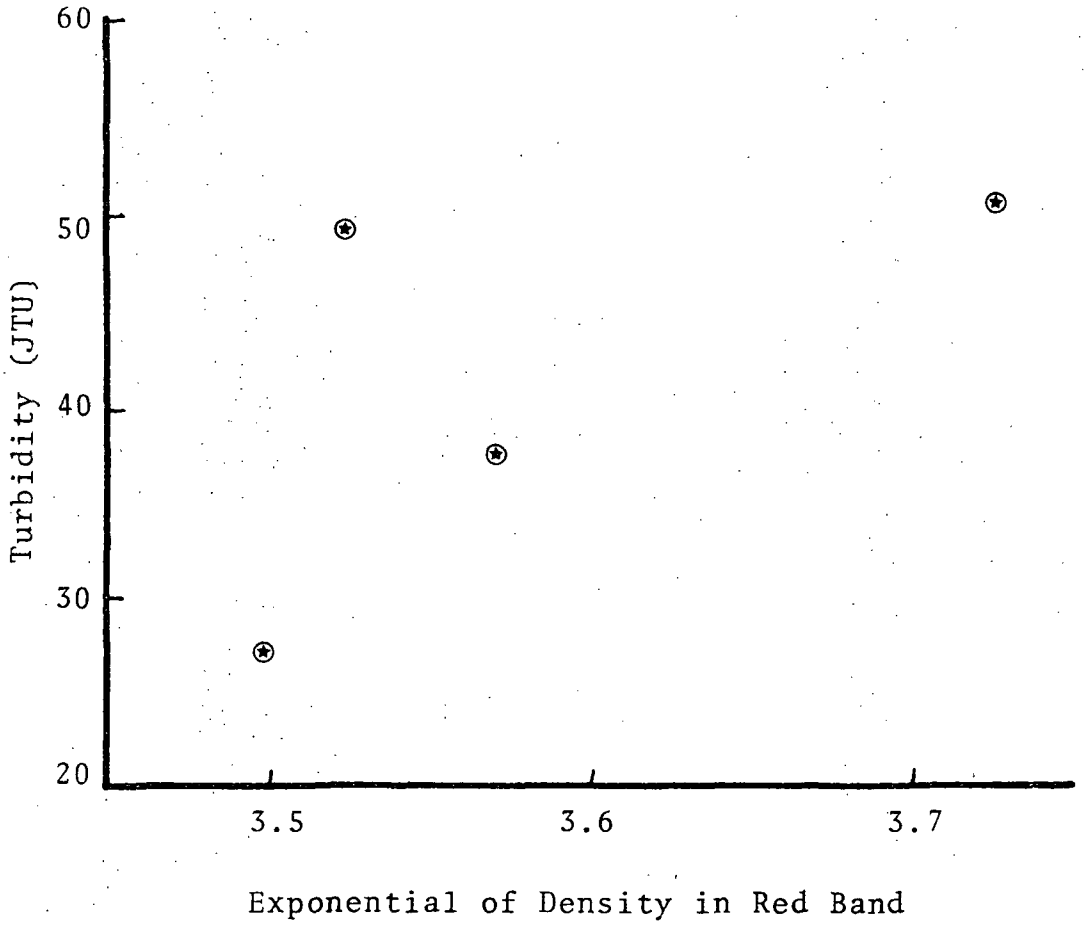


Fig. A-13 Photographic Data Scatter Diagram

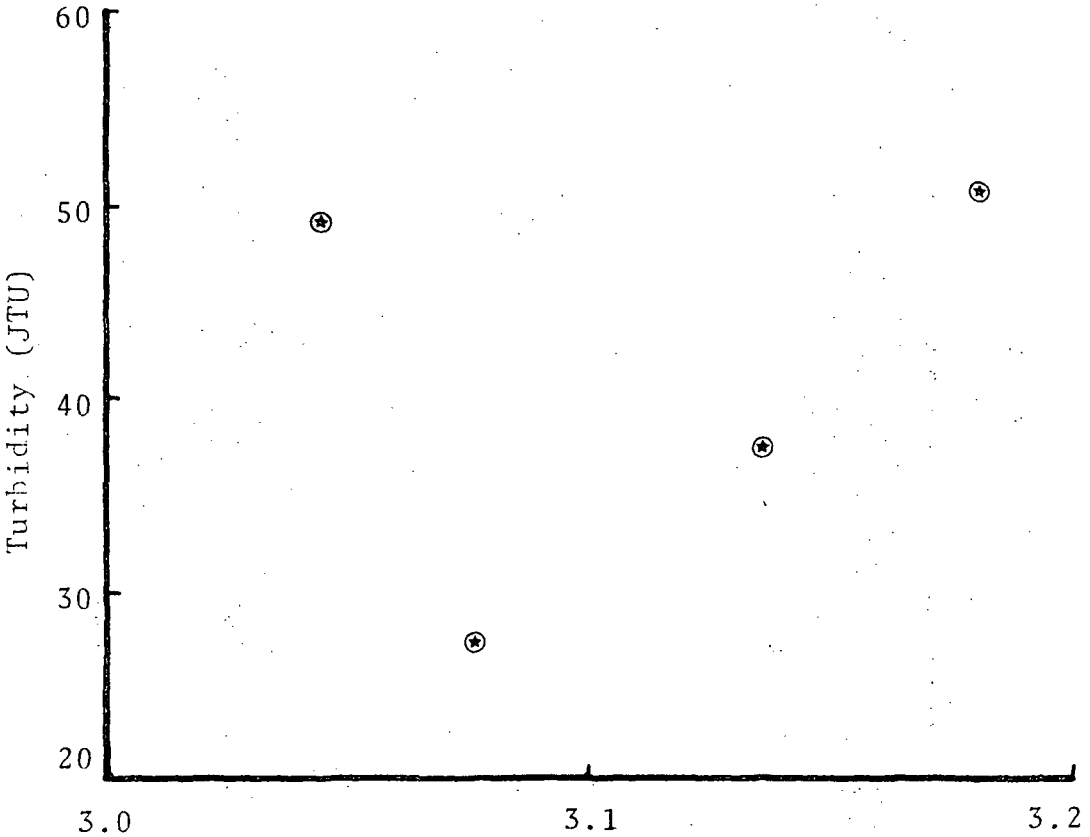


Fig. A-14 Photographic Data Scatter Diagram

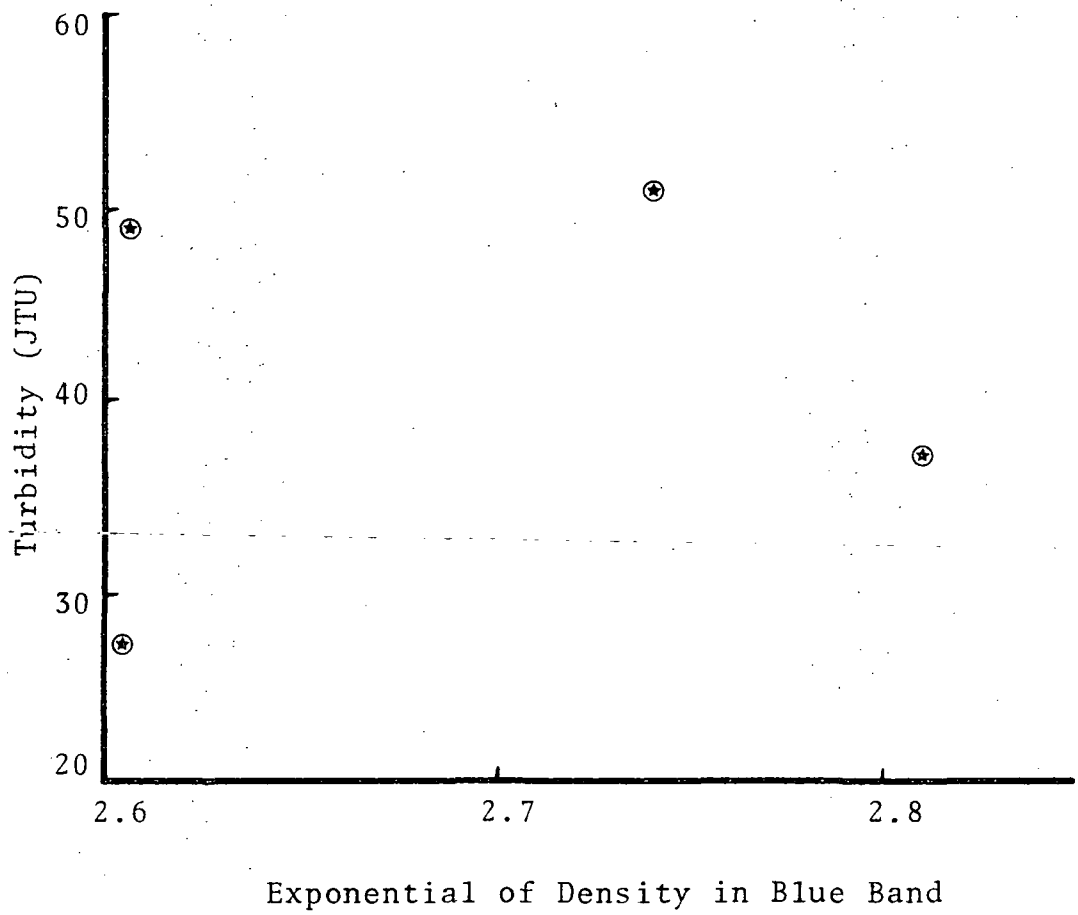


Fig. A-15 Photographic Data Scatter Diagram

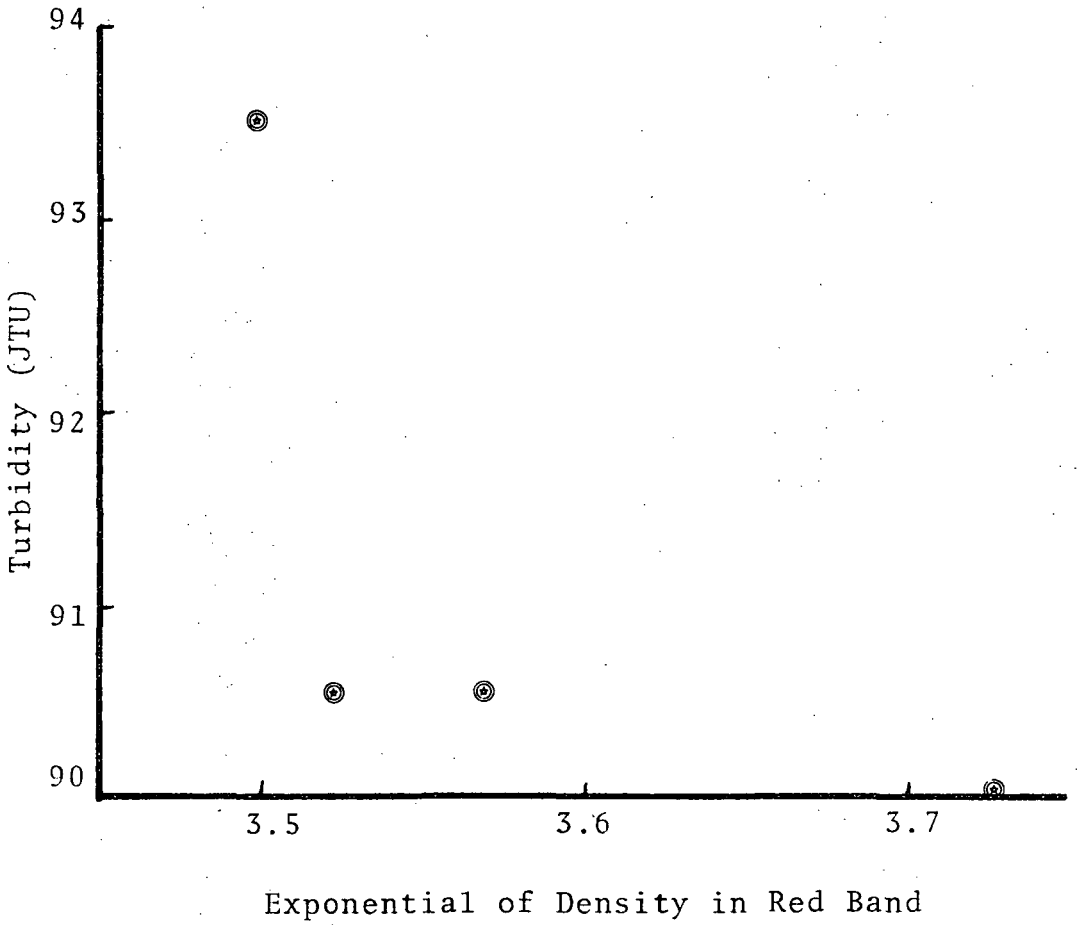


Fig. A-16 Photographic Data Scatter Diagram

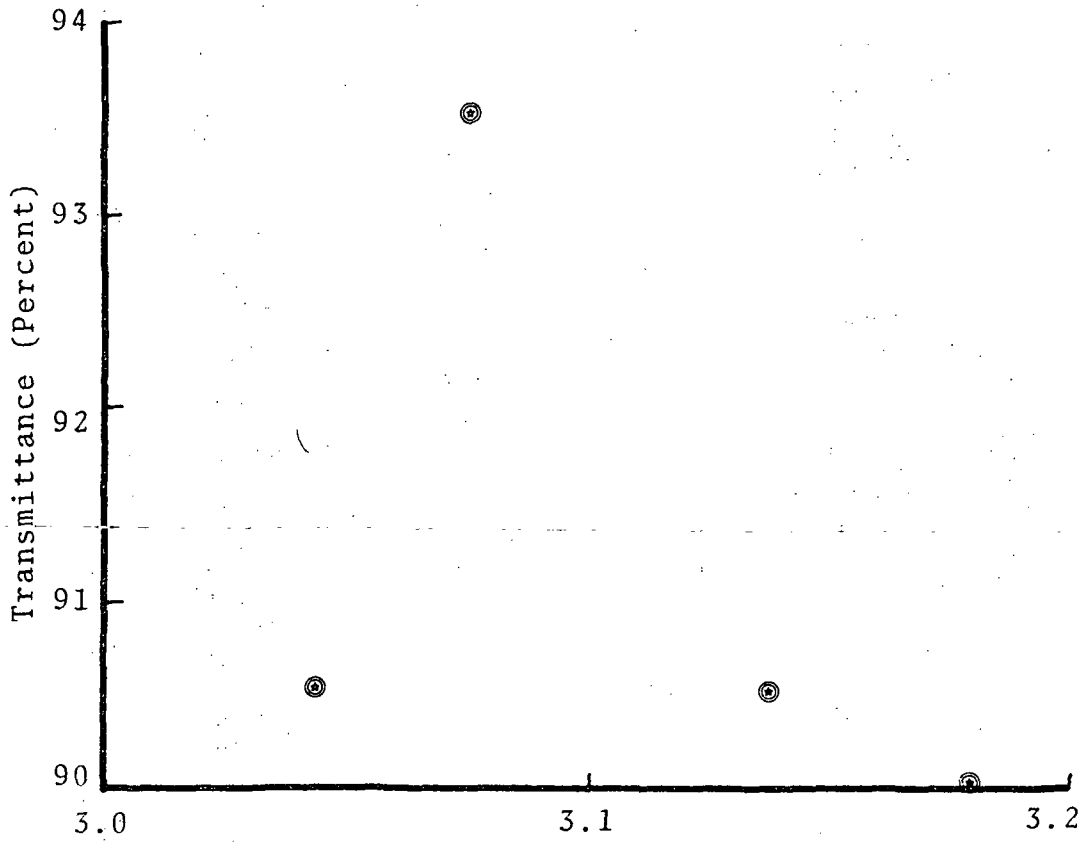


Fig. A-17 Photographic Data Scatter Diagram

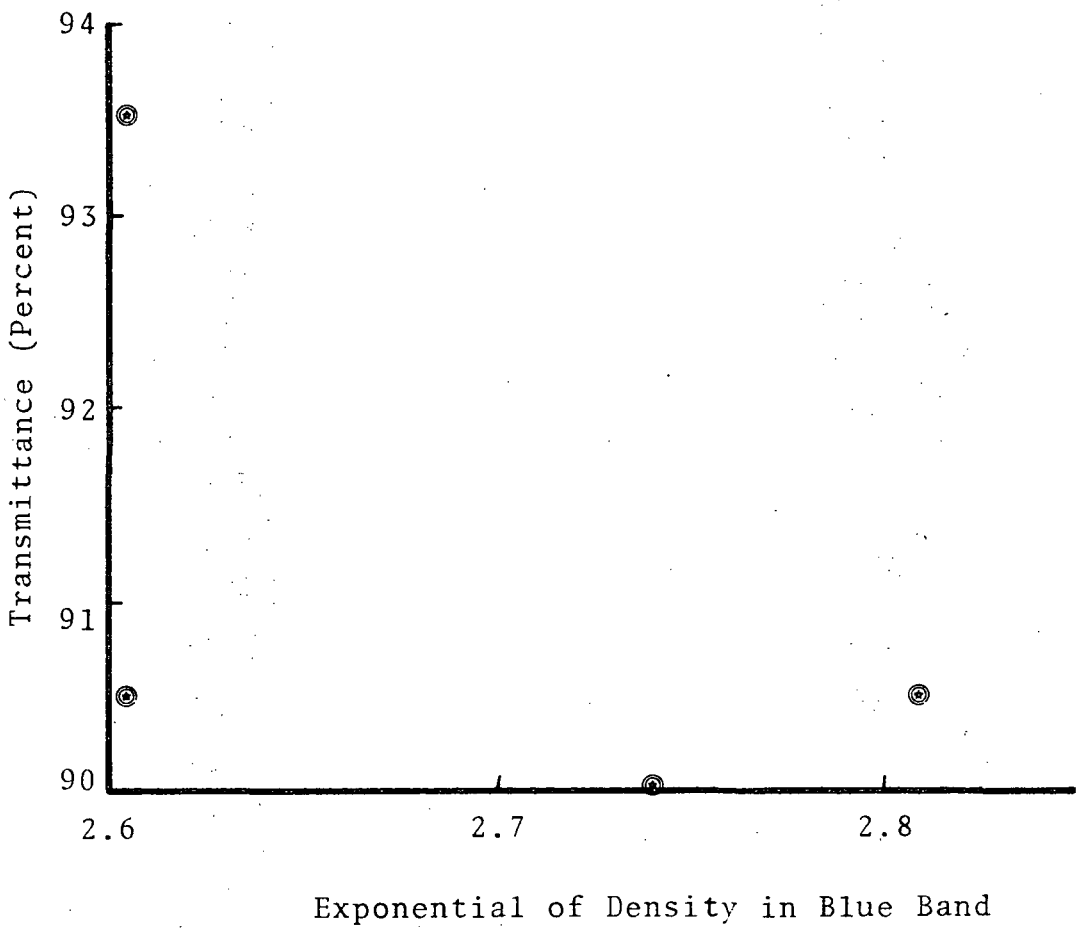


Fig. A-18 Photographic Data Scatter Diagram

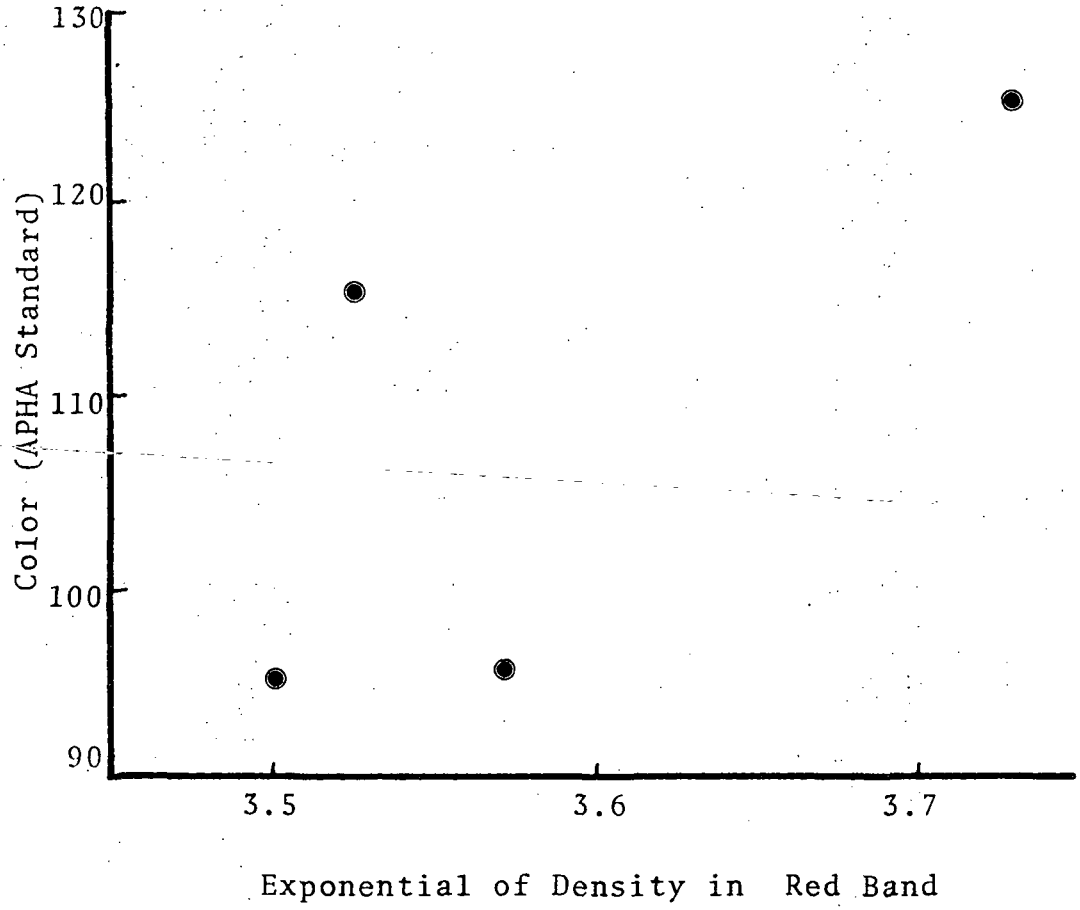


Fig. A-19 Photographic Data Scatter Diagram

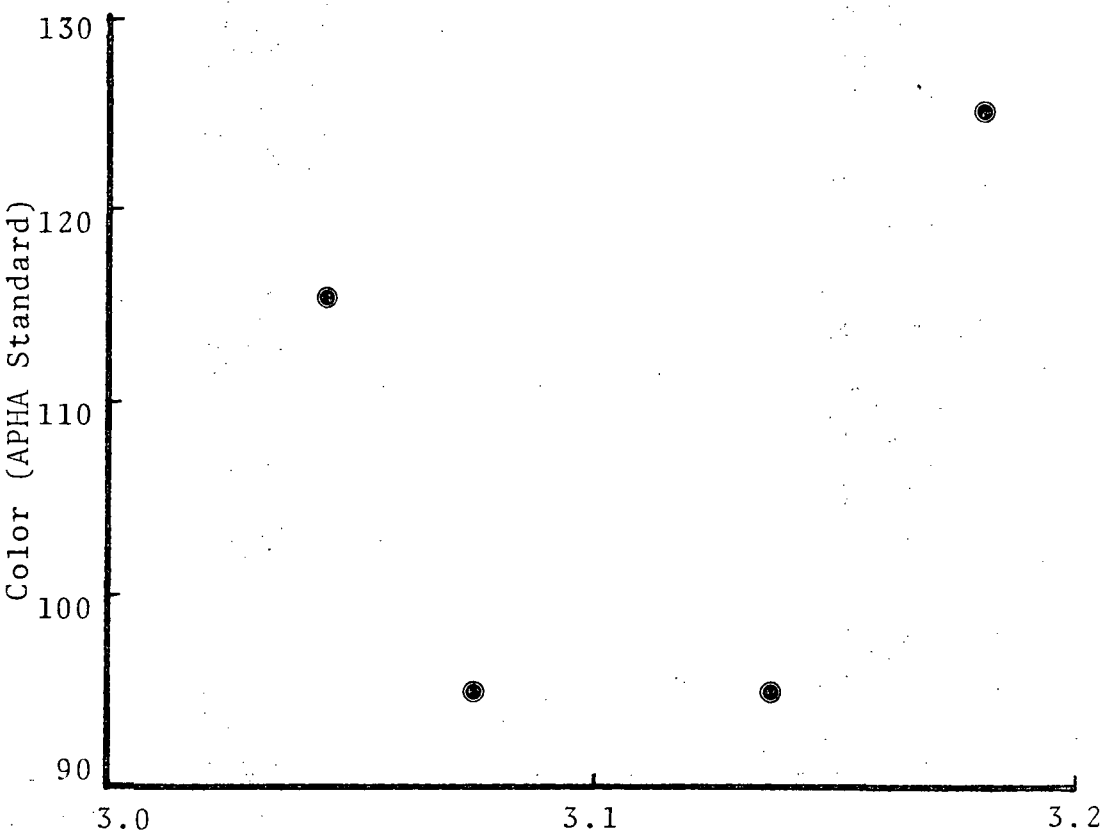


Fig. A-20 Photographic Data Scatter Diagram

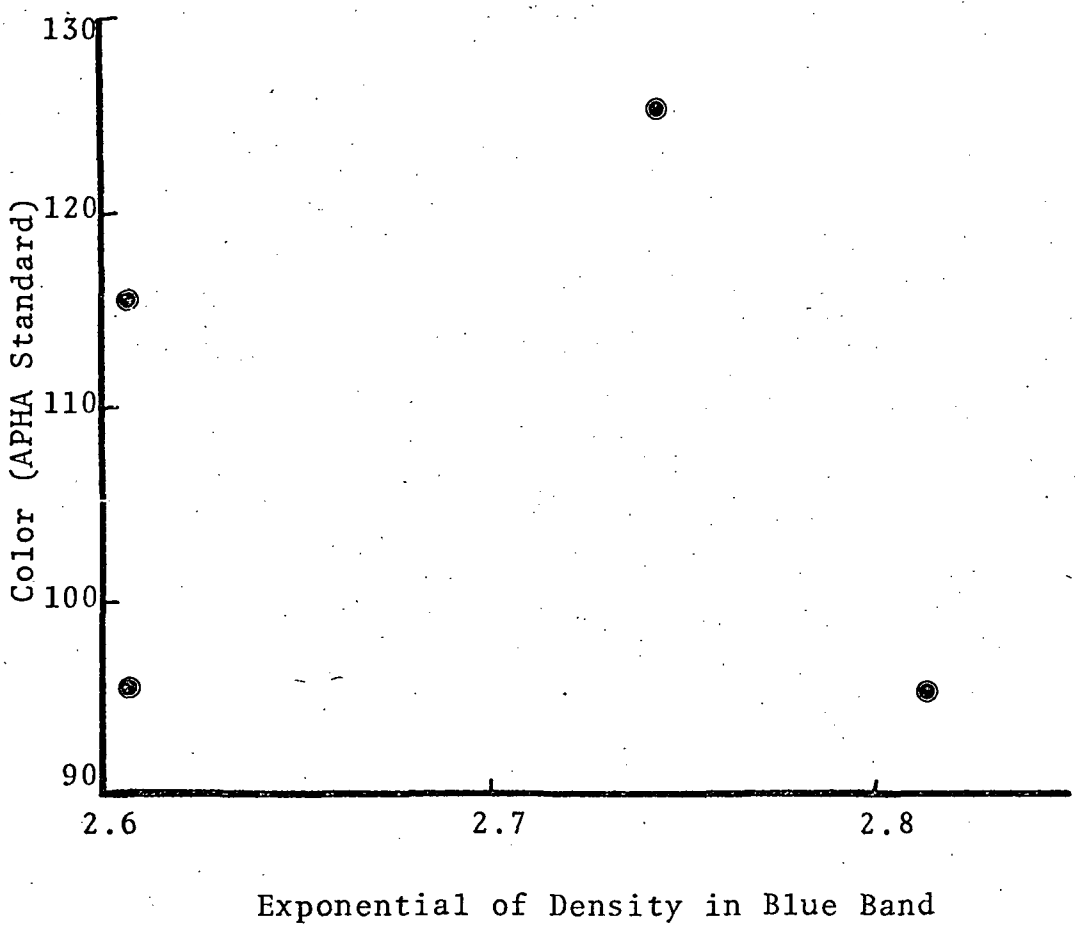


Fig. A-21 Photographic Data Scatter Diagram

The REMOTE SENSING CENTER was established by authority of the Board of Directors of the Texas A&M University System on February 27, 1968. The CENTER is a consortium of four colleges of the University; Agriculture, Engineering, Geosciences, and Science. This unique organization concentrates on the development and utilization of remote sensing techniques and technology for a broad range of applications to the betterment of mankind.